UNIVERSITÉ DE SHERBROOKE
Faculté de génie
Département de génie électrique et de génie informatique

REHAUSSEMENT DE LA PAROLE COMBINANT LE DOMAINE SPECTRAL ET LE DOMAINE DES MODULATIONS DU SPECTRE

Mémoire de maitrise
Spécialité : génie électrique

Julien Bosco

Sherbrooke (Québec) Canada

Mai 2018

MEMBRES DU JURY

Éric PLOURDE

Directeur

Jean ROUAT

Évaluateur

Sébastien ROY

Évaluateur

RÉSUMÉ

Ce projet de recherche propose une technique de rehaussement de la parole basée sur le domaine spectral et le domaine des modulations du spectre. Dans cette approche, un signal de parole bruité est rehaussé dans le domaine spectral en utilisant un estimateur d'erreur quadratique moyenne minimale (EQMM) du spectre et dans le domaine des modulations du spectre avec un estimateur EQMM des modulations du spectre. Les résultats de chaque estimateur sont combinés, à l'aide d'une fonction basée sur le rapport signal à bruit (RSB) *a priori* du signal de parole bruité, pour obtenir le signal de parole rehaussé. Des résultats comparatifs à partir du score PESQ (*Perceptual Evaluation of Speech Quality*), du RSB par segments temporels et du rapport signal à distorsion (RSD) sont présentés afin de valider la performance de la technique proposée par rapport aux techniques existantes. La technique proposée donne une réduction de bruit supérieure par rapport aux techniques présentées, mais a comme conséquence d'introduire de la distorsion dans le signal de parole rehaussé.

Mots-clés : rehaussement de la parole, domaine spectral, domaine modulatoire, estimation de bruit

À ma fille Lilas, qui me fait commencer une nouvelle aventure à la fin de celle de ma maitrise.

REMERCIEMENTS

Un gros merci à mon directeur de recherche Éric Plourde, pour son aide et sa patience durant toute ma maitrise. Un merci aussi à toute l'équipe de NECOTIS pour leur assistance et leur bonne humeur !

TABLE DES MATIÈRES

1 Introduction **1**
1.1 Mise en contexte et problématique . . . 1
1.1.1 Bref historique du traitement de la parole . . . 1
1.1.2 Rehaussement de la parole . . . 3
1.2 Définition du projet de recherche . . . 3
1.3 Objectifs du projet de recherche . . . 4
1.3.1 Objectif principal général . . . 4
1.3.2 Objectifs principaux spécifiques . . . 4
1.4 Contributions originales . . . 4
1.5 Plan du document . . . 5

2 État de l'art **7**
2.1 Transformée de Fourier . . . 7
2.2 Structure AMS . . . 8
2.2.1 Analyse . . . 8
2.2.2 Synthèse . . . 10
2.3 Techniques de rehaussement de la parole . . . 10
2.3.1 Rehaussement par soustraction spectrale . . . 10
2.3.2 Rehaussement par estimateur spectral EQMM . . . 12
2.3.3 Rehaussement par soustraction spectrale dans le domaine des modulations du spectre . . . 14
2.3.4 Rehaussement par estimateur EQMM des modulations du spectre . 16
2.3.5 Rehaussement avec la technique de Fusion . . . 18
2.4 Estimation du bruit par estimation des statistiques minimales du bruit . . 20
2.4.1 Principes d'estimation des statistiques minimales du bruit . . . 20
2.4.2 Optimisation du facteur de correction en fonction de l'erreur d'estimation . . . 22
2.4.3 Estimateur de bruit non biaisé de statistique minimale . . . 23

3 Méthodologie **27**
3.1 Composition de la banque de segments bruités . . . 28
3.1.1 Base de données de segment de paroles . . . 28
3.1.2 Banque de bruits . . . 28
3.1.3 Mixage des segments de parole avec les segments de bruit . . . 29
3.2 Changement de cadence de 8 kHz à 16 kHz . . . 29
3.2.1 Déploiement des RTC et réseaux cellulaires à large bande . . . 29
3.2.2 Modification des algorithmes de rehaussement . . . 30
3.3 Développement de la fonction de combinaison de l'algorithme EQMM double 38

4 Résultats expérimentaux **41**
4.1 Mesure des performances d'un algorithme de rehaussement de la parole . . 41

4.1.1 Mesure de qualité de la parole pour un système de télécommunication : PESQ 41
4.1.2 RSB par segments temporels 41
4.1.3 Mesure de distorsion : Rapport signal à distorsion (RSD) 42
4.2 Résultats et discussions 42
4.2.1 Bruits stationnaires 42
4.2.2 Bruits non stationnaires 46

5 Conclusion **53**

LISTE DES RÉFÉRENCES **55**

LISTE DES FIGURES

2.1 Fonction de poids, $\Psi[\omega[l]]$ 18

3.1 Effet de la durée d'une trame spectrale sur le score PESQ pour l'algorithme EQMM AS. Une durée de trame d'environ 15 ms permet d'avoir le meilleur score. 30
3.2 Effet du facteur de lissage α sur le score PESQ pour l'algorithme EQMM AS. Une valeur de 0.99 permet d'avoir le meilleur score PESQ. 31
3.3 Effet du seuil minimum $\gamma_{\min}$ pour le RSB *a priori* sur le score PESQ pour l'algorithme EQMM AS. La variation du seuil minimum n'a aucune influence sur le score PESQ. 32
3.4 Effet de la durée d'une trame spectrale sur le score PESQ pour l'algorithme EQMM MMS. Une durée de trame de 15 ms permet d'avoir le meilleur score PESQ. 33
3.5 Effet de la durée d'une trame modulatoire sur le score PESQ pour l'algorithme EQMM MMS. Une durée de trame de 256 ms permet d'avoir le meilleur score PESQ. 34
3.6 Effet du facteur de lissage α sur le score PESQ pour l'algorithme EQMM MMS. Une valeur de 0.99 permet d'obtenir le meilleur score PESQ. 34
3.7 Effet du seuil minimum $\gamma_{\min}$ pour le RSB *a priori* sur le score PESQ pour l'algorithme EQMM MMS. Un seuil de 0.005 permet d'obtenir le score PESQ maximum à faible RSB. 35
3.8 Effet de la durée d'une trame spectrale sur le score PESQ pour l'algorithme SMS. Une durée de trame de 12 ms permet d'obtenir le meilleur score PESQ. 36
3.9 Effet de la durée d'une trame modulatoire sur le score PESQ pour l'algorithme SMS. Une durée de trame de 256 ms permet d'obtenir le meilleur score PESQ. 37
3.10 Effet du facteur ρ sur le score PESQ pour l'algorithme SMS. La variation du facteur n'a pas assez d'incidence sur le score PESQ. La valeur de 3 proposée par défaut est utilisée. 37
3.11 Effet du facteur β sur le score PESQ pour l'algorithme SMS. La valeur de 0.01 est utilisée étant donné le léger gain à 15 dB. 38

4.1 RSB segmental (dB) pour de la parole bruitée avec un BABG. L'algorithme EQMM double a le meilleur RSB segmental pour tout RSB. 43
4.2 RSD (dB) pour de la parole bruitée avec un BABG. L'algorithme EQMM double présente plus de distorsion que les algorithmes à un seul domaine. . 44
4.3 RSB segmental (dB) pour de la parole bruitée avec un BAR. L'algorithme EQMM double a le meilleur RSB segmental pour tout RSB. 45
4.4 RSD (dB) pour de la parole bruitée avec un BAR. L'algorithme EQMM double présente plus de distorsion que les algorithmes à un seul domaine. . 46

4.5 RSB segmental (dB) pour de la parole bruitée avec bruit additif de conversations. L'algorithme EQMM double a le meilleur RSB segmental pour tout RSB. 47
4.6 RSD (dB) pour de la parole bruitée avec bruit additif de conversations. L'algorithme EQMM double présente plus de distorsion que les algorithmes à un seul domaine. 47
4.7 RSB segmental (dB) pour de la parole bruitée avec un bruit additif d'usine. L'algorithme EQMM double a le meilleur RSB segmental pour tout RSB. . 48
4.8 RSD (dB) pour de la parole bruitée avec un bruit additif d'usine. L'algorithme EQMM double présente plus de distorsion que les algorithmes à un seul domaine. 49
4.9 RSB segmental (dB) pour de la parole bruitée avec un bruit additif de voiture. L'algorithme EQMM double a le meilleur RSB segmental pour tout RSB. 50
4.10 RSD (dB) pour de la parole bruitée avec un bruit additif de voiture. L'algorithme EQMM double présente plus de distorsion que les algorithmes à un seul domaine. 51

LISTE DES TABLEAUX

2.1 Score PESQ avec un bruit additif blanc gaussien (BABG). 19

4.1 Score PESQ moyen avec bruit additif blanc gaussien (BABG). 43
4.2 Score PESQ moyen avec bruit additif rose (BAR). 44
4.3 Score PESQ moyen avec bruit additif de conversation. 46
4.4 Score PESQ moyen avec bruit additif d'usine. 48
4.5 Score PESQ moyen avec bruit additif d'usine. 50

LISTE DES ACRONYMES

Acronyme	Définition
AMR-WB	*Adaptive Multi-Rate Wideband*
AMS	Analyse-Modification-Synthèse
BABG	Bruit additif blanc gaussien
BAR	Bruit additif rose
DAV	Détecteur d'activité vocale
DSP	Densité spectrale de puissance
EQMM	Erreur quadratique minimum moyenne
EQMM AS	EQMM de l'amplitude spectrale
EQMM MMS	EQMM du module des modulations du spectre
FDP	Fonction de densité de probabilité
HMM	*Hidden Markov Model*
ITU-T	*Internation Telecommunication Union - Telecommunication*
MMSE	*Minimum Mean Square Error*
MOS	*Mean Opinion Score*
PESQ	*Perceptual Evaluation of Speech Quality*
RSB	Rapport signal à bruit
RSD	Rapport signal à distorsion
RTC	Réseau téléphonique commuté
SMS	Soustraction des modulations du spectre
TF	Transformée de Fourier
TFD	Transformée de Fourier discrète
TFDI	Transformée de Fourier discrète inverse
TFR	Transformée de Fourier rapide
TFSD	Transformée de Fourier de signaux discret
VoLTE	*Voice over Long Term Evolution*

CHAPITRE 1

Introduction

1.1 Mise en contexte et problématique

1.1.1 Bref historique du traitement de la parole

La première tentative de traitement de la parole remonte en 1769 quand Kratzenstein, un médecin physicien ingénieur allemand, fabriqua un système composé de cavités résonnantes qui permettait de reproduire les voyelles /a/,/e/,/i/,/o/,/y/ en faisant vibrer un roseau. Durant la même période, Wolfgang von Kempelen conçut un synthétiseur vocal mécanique générant voyelles, consonnes et mélange simple celles-ci. Ces travaux sont considérés comme le début du traitement de la parole [31]. Il faut attendre les années 1920 et 1930 pour voir apparaitre les premiers synthétiseurs électriques. L'une des plus grandes avancées durant ces années-là vient de Homer Dudley, ingénieur chez Bell Labs. Il est le premier à montrer la parole comme une porteuse d'information [30], en analogie avec le principe de porteuse en transmission radio. Durant sa carrière, il développa les deux premiers systèmes de traitement de la parole commerciaux : le Voder et le Vocoder. Le premier fut un synthétiseur de parole tandis que le deuxième servait à compresser la parole. Les deux produits n'ont jamais réussi à devenir populaires étant donné la grande difficulté à produire ou compresser la parole tout en conservant son intelligibilité. Le Vocoder a été, par contre, utilisé par les militaires durant la Seconde Guerre mondiale, car il avait la capacité de pouvoir crypter les signaux de parole et parce qu'une mauvaise qualité audio était tolérée pour les communications [4].

Bien que les travaux de Dudley n'ont pas été une réussite au point de vue commercial, ils ont été la base pour la majorité des recherches sur le traitement de la parole qui ont eu lieu depuis. Il est le pionnier des technologies utilisées pour la compression, la reconnaissance et la synthèse de la parole qui sont introduites dans les sous-sections suivantes.

Compression de la parole

Les premières techniques de compression proposées ont visé une réduction de la bande passante du signal acoustique [4]. En effet, un signal téléphonique filaire, par exemple,

se situe dans la bande de fréquence de 0 Hz à 3.4 kHz, bien que l'oreille puisse entendre jusqu'à environ 20 kHz. Ceci est possible puisque la majeure partie de l'énergie d'un signal de parole est concentrée dans les fréquences inférieures à 3.4 kHz. Il est donc possible d'avoir une bonne compréhension en ne conservant que les fréquences en bas de 3.4 kHz. La possibilité de réduire cette bande plus bas que 3.4 kHz permet donc de réduire la bande passante utilisée. Une bande passante réduite permet d'utiliser une porteuse de plus basse fréquence, augmentant la portée de communication, mais avec une plus grande perte de qualité du signal.

Par la suite, les découvertes des années 40 et 50 sur la théorie de l'information ont démontré qu'il était plus intéressant de réduire le débit d'information plutôt que la bande passante du signal. Grâce à l'avènement des systèmes numériques, de nouvelles techniques de compression ont permis de réduire le débit d'information tout en conservant une qualité audio acceptable. Pour un signal téléphonique typique échantillonné à 8 kHz quantifié sur 8 bits, un débit de 64 kbit/s est nécessaire pour transmettre sans perte de qualité. Il est possible avec les techniques d'aujourd'hui de réduire ce débit à 13 kbit/s avec une très faible perte de qualité. Pour les communications téléphoniques, le débit peut être réduit aussi bas que 2.4 kbit/s tout en conservant une bonne intelligibilité de la parole. Des tentatives ont été faites pour réduire le débit à 300 bits/s pour les communications militaires, mais il est alors presque impossible de pouvoir comprendre la conversation. [4]

Reconnaissance vocale

Une autre application du traitement de la parole est la reconnaissance vocale. Les premières techniques de reconnaissance consistaient à simplement reconnaître des mots dans un vocabulaire limité (une centaine de mots possibles). C'est durant les années 1980, avec, entre autre, l'introduction de modèles de Markov caché (*HMM* en anglais) que la reconnaissance vocale a pu s'améliorer en permettant de reconnaître des phrases complètes [4]. Une des plus grandes difficultés de la reconnaissance vocale est de pouvoir être assez généralisée pour comprendre une grande quantité de personnes ayant des accents, des expressions et même des dialectes différents et de pouvoir les différencier. La reconnaissance vocale est aussi très sensible à la présence de bruit, ce qui vient diminuer grandement la performance des algorithmes de reconnaissance vocale.

Synthèse vocale

Une autre grande application du traitement de la parole est de pouvoir générer de la parole. Les premières tentatives de synthèse ont consisté à convertir un texte en parole. Le système contient alors une série de phonèmes reliés à des syllabes ou des séries de syllabes

et au fur et à mesure que le texte est traité, les phonèmes sont produits un à la suite de l'autre avec les silences nécessaires pour produire un signal de parole compréhensible.

La synthèse s'est beaucoup améliorée aujourd'hui, car ces systèmes contiennent non plus seulement des phonèmes, mais aussi des morceaux complets de parole qui sont peu ou pas du tout modifiés. Couplée avec les nouvelles techniques d'intelligence artificielle, la synthèse vocale permet une conversation avec une machine sans devoir disposer de textes écrits au préalable. La parole est générée automatiquement selon l'entrée donnée au système. [4]

1.1.2 Rehaussement de la parole

Pour que ces différentes techniques de traitement de la parole puissent fonctionner correctement, il est important de disposer en entrée d'un signal de parole clair et intelligible. Or, dans la pratique, la parole est acquise en présence de bruits, qui viennent dégrader la performance de ces techniques et qui peut même les rendre non fonctionnelles. Il est donc nécessaire de devoir rehausser cette parole bruitée pour que ces techniques puissent être utilisées de manière efficace. Ce rehaussement consiste à retirer l'énergie du bruit pour atteindre l'intelligibilité du signal de parole propre. Pour ce faire, plusieurs méthodes de rehaussement de la parole ont été développées afin de rendre les techniques de traitement de la parole fonctionnelles et performantes. Les premières techniques développées consistaient à supprimer le bruit en utilisant la soustraction spectrale [6]. Des techniques plus complexes, en utilisant des estimateurs statistiques sur le spectre [11] et plus récemment sur la modulation du spectre [28] [29], ont permis d'améliorer la capacité à réduire le bruit dans un segment de parole.

1.2 Définition du projet de recherche

Le projet de recherche est défini selon la question de recherche suivante : *est-il possible de développer un algorithme de rehaussement de la parole utilisant le domaine spectral et les modulations du spectre afin d'obtenir une performance supérieure aux techniques actuelles pour les systèmes de traitement de la parole moderne* ? Ce sujet s'appuie sur de récentes découvertes sur l'efficacité de techniques de rehaussement utilisant le domaine spectral et les modulations du spectre [28] [29].

1.3 Objectifs du projet de recherche

1.3.1 Objectif principal général

L'objectif principal est de pouvoir développer une technique de préfiltrage à un système de traitement de la parole pour réduire au maximum l'impact du bruit et permettre de maximiser la performance de ce système. Malgré le fait qu'il existe beaucoup de techniques de rehaussement de la parole, il est encore difficile de disposer d'un rehaussement acceptable lorsque le RSB est faible.

1.3.2 Objectifs principaux spécifiques

Il y a deux objectifs spécifiques à atteindre dans ce projet.

Développer un algorithme de rehaussement de la parole utilisant les domaines spectraux et des modulations du spectre

Le but est de réussir à tirer avantage du domaine des modulations du spectre afin d'obtenir une meilleure performance que les techniques utilisant uniquement le domaine spectral.

Évaluer les performances des algorithmes utilisant la modulation de spectre avec un changement de cadence de 8 à 16 kHz.

Avec l'arrivée de système de télécommunication comme la VoTLE qui utilise une fréquence d'échantillonnage de 16 kHz, il est important que les techniques de rehaussement puissent fonctionner correctement à cette cadence et puissent converser une bonne performance.

1.4 Contributions originales

Dans ce travail de recherche, nous proposons un algorithme qui combine l'utilisation simultanée du domaine spectral et du domaine des modulations du spectre. Cet algorithme utilise un estimateur EQMM pour le rehaussement dans chaque domaine, puis reconstruit le signal de parole rehaussé à partir d'une combinaison de ces deux domaines, en fonction du RSB *a priori* du signal de parole bruité. Les résultats démontrent de meilleurs résultats en termes de mesure de performance objective de type PESQ, mais au prix d'une augmentation de la distorsion. Nous démontrons également que pour une fréquence d'échantillonnage de 16 kHz, l'algorithme proposé conserve des performances semblables aux algorithmes utilisés avec une fréquence d'échantillonnage de 8 kHz, confirmant son fonctionnement pour une fréquence d'échantillonnage de 16 kHz.

Les travaux de cette maitrise ont fait l'objet d'un article qui a été accepté dans le cadre de la *2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE)* [7]. Les travaux ont aussi été présentés durant cette conférence qui a eu lieu du 30 avril au 3 mai 2017.

1.5 Plan du document

Le document est structuré de la façon suivante. Le chapitre 2 présente une revue de littérature du modèle AMS (Analyse-Modification-Synthèse) utilisant la transformée de Fourier (TF) ainsi que les différentes techniques actuelles de rehaussement de la parole utilisant entre autres la modulation du spectre. Au chapitre 3, les détails pour l'implémentation de l'algorithme proposé de rehaussement utilisant les deux domaines spectral et des modulations du spectre sont fournis. Le chapitre 4 fait état des tests et résultats des comparaisons faites entre la technique proposée et celles présentées dans l'état de l'art. Finalement, le chapitre 5 récapitule les travaux de la recherche et présente une ouverture sur les futures améliorations possibles.

CHAPITRE 2

État de l'art

Les différentes techniques de rehaussement présentées dans ce mémoire se basent sur un signal de parole corrompu par un bruit additif et non corrélé tel que

$$x[n] = s[n] + d[n] \tag{2.1}$$

où n représente l'index temporel discret et $x[n]$, $s[n]$ et $d[n]$ représentent respectivement la parole bruitée, le signal de parole propre et le bruit.

Une des approches classiques de rehaussement de la parole consiste à transformer le signal du domaine temporel à un domaine choisi, par exemple le domaine spectral à l'aide de la transformée de Fourier (TF), et d'effectuer ensuite le rehaussement dans ce domaine plutôt que directement dans le domaine temporel. Une fois le rehaussement effectué, le signal est ensuite retransformé dans le domaine temporel. On l'appelle approche par analyse, modification et synthèse (AMS). C'est cette approche qui sera privilégiée dans ce mémoire.

Dans ce chapitre, nous présenterons la transformée de Fourier discrète (TFD), nécessaire pour les parties analyse et synthèse de l'approche AMS. Ensuite l'approche AMS est présentée. Les techniques de rehaussement de la parole suivantes sont présentées : rehaussement par soustraction spectrale, par estimeur EQMM spectral, par soustraction spectrale dans le domaine des modulations du spectre, par estimateur EQMM des modulations du spectre et avec la technique de Fusion. Finalement, l'estimateur de bruit par estimation des statistiques minimales est présenté.

2.1 Transformée de Fourier

Puisque les signaux de parole sont échantillonnés, la transformée de Fourier pertinentes permettant d'obtenir une représentation dans le domaine spectral du signal d'origine est la transformée de Fourier discrète (TFD). Elle est définie par

$$X[k] = \sum_{n=0}^{N-1} x[n] \cdot e^{-j\frac{2\pi nk}{N}} \tag{2.2}$$

où k et N correspondent respectivement à l'index fréquentiel et la taille du signal $x[n]$. Comme les signaux de parole peuvent être considérés comme stationnaire sur des intervalles dont la durée est comprise entre 20 à 30ms. [5], il est nécessaire de séparer le signal en petites portions de 20 à 30 ms à l'aide de fenêtres si l'on désire appliquer la TFD. Cet aspect est présenté dans la section suivante dans le contexte d'une structure AMS.

2.2 Structure AMS

La structure AMS sépare le rehaussement en trois parties : l'analyse, la modification et la synthèse. La première partie permet de séparer le signal en trame et de changer de domaine, passant du domaine temporel au domaine spectral en utilisant la TFD. La seconde partie permet d'appliquer la technique de rehaussement sur le signal dans le domaine spectral. La dernière partie utilise la TFDI (transformée de Fourier discrète inverse) afin de repasser du domaine spectral au domaine temporel.

2.2.1 Analyse

L'analyse permet de séparer le signal en trame et de convertir le signal du domaine temporel au domaine spectral. L'équation de la TFD utilisée pour cette conversion est

$$X[\eta, l] = \sum_{n=0}^{N-1} x[n + lZ] w_a[n] e^{-j\frac{2\pi\eta n}{N}} \tag{2.3}$$

où η, l, N, $w_a[\cdot]$ et Z correspondent respectivement à l'index fréquentiel, l'index de trame, la longueur d'une trame, la fenêtre d'analyse et le décalage pour la superposition des trames. On obtient donc une matrice contenant l'information spectrale de chaque trame. Les différents algorithmes de rehaussement étudiés dans ce mémoire s'appliquent au module du spectre plutôt qu'au spectre directement. Le spectre $X[\cdot]$ peut être représenté sous sa forme polaire de la manière suivante :

$$X[\eta, l] = \Big|X[\eta, l]\Big| e^{j\angle X[\eta, l]} \tag{2.4}$$

où $\Big|X[\eta, l]\Big|$ corresponds au module du spectre et $\angle X[\eta, l]$ à sa phase. Comme la phase a peu d'incidence sur le rehaussement de la parole [27], le rehaussement se fera uniquement sur le module du spectre. Lors de la synthèse, la phase bruitée sera utilisée pour la reconstruction.

Fonctions de fenêtrage

Le signal de parole est fortement non-stationnaire. Afin de pouvoir utiliser des opérateurs, tel qu'une transformée de Fourier (qui suppose la stationnarité du signal), il est nécessaire de séparer le signal en trames de telle sorte que chaque trame puisse être considérée comme étant un signal stationnaire. Les fonctions de fenêtrage sont des fonctions de poids qui s'appliquent directement sur chaque trame. Il existe plusieurs types de fenêtres ayant chacune des propriétés différentes.

Fonction rectangulaire La fonction rectangulaire conserve l'amplitude du signal intacte à l'intérieur d'un nombre donné d'échantillons et nul part ailleurs. Son équation est

$$w_{\text{rect}}[n] = 1, \quad n = 0, 1, ..., N-1. \tag{2.5}$$

Ce type de fenêtrage est généralement évité puisqu'il ne permet pas d'atténuer les bornes de la trame. En effet, afin d'éviter certains artéfacts dus au traitement effectué sur les trames, il est préférable que la fonction de fenêtrage atténue les bornes de la trame.

Fonction de Hann La fonction de Hann fait partie de la famille des fenêtres de type cosinus. Les avantages de cette famille est la facilité d'obtenir les propriétés spectrales de façon analytique et qu'elle soit nulle à ces extrémités. Son équation pour un signal discret est donnée par

$$w_{\text{Hann}}[n] = 0.5 + 0.5\cos(2n\pi/N). \tag{2.6}$$

Cette fenêtre débute et termine à zéro, ce qui permet de couper toute discontinuité. D'un point de vue spectral, elle n'a pas une atténuation aussi rapide que la fenêtre de Hamming proche du premier lobe, mais elle a une meilleure atténuation plus on s'éloigne du premier lobe [14]

Fonction de Hamming La fonction de Hamming est similaire à celle de Hann. Son équation est donnée par

$$w_{\text{Hamming}}[n] = 0.54 - 0.46\cos[\frac{2n}{N}\pi] \quad \text{n} = 0, 1, 2, \cdots, N-1 \tag{2.7}$$

La fenêtre n'est pas complètement nulle à ses extrémités, ce qui cause de petites discontinuités. Cette fenêtre à une meilleure atténuation que Hann autour du premier lobe, mais devient plus faible aux lobes subséquents [14]

2.2.2 Synthèse

La synthèse reprend les étapes de l'analyse à l'inverse pour reconstruire le signal de parole à partir du spectre rehaussé. Pour chacune des trames l, le spectre rehaussé, $\hat{S}[\eta, l]$, est obtenu en combinant le module du spectre rehaussé avec la phase bruitée :

$$\hat{S}[\eta, l] = \left|\hat{S}[\eta, l]\right| * e^{j\angle X[\eta,l]} \tag{2.8}$$

On applique ensuite la transformée de Fourier inverse au spectre rehaussé :

$$\hat{s}_l[n] = \frac{1}{N}\Big\{w_s[n] \sum_{\eta=0}^{N-1} \hat{S}[\eta, l] e^{j\frac{2\pi\eta n}{N}}\Big\} \tag{2.9}$$

où $w_s[n]$ est la fenêtre de synthèse. On recombine ensuite les trames à l'aide de la technique de recouvrement et d'addition (*overlap and add)* [26] :

$$\hat{s}[n] = \sum_{l=0}^{N_t-1} \hat{s}_l[n - lZ] \tag{2.10}$$

où N_t correspond au nombre de trames créées dans la partie analyse.

2.3 Techniques de rehaussement de la parole

Dans cette section, les techniques de rehaussement suivantes sont présentées : rehaussement par soustraction spectrale, par estimeur EQMM spectral, par soustraction spectrale dans le domaine des modulations du spectre, par estimateur EQMM des modulations du spectre et avec la technique de Fusion.

2.3.1 Rehaussement par soustraction spectrale

Une des premières techniques de rehaussement dans le domaine spectral ayant été proposée dans la littérature est le rehaussement par soustraction spectrale. Le but est d'estimer le module du spectre du signal de parole débruité en soustrayant le module du spectre du bruit au module du spectre du signal de parole bruité [6]. Le spectre du bruit pourra, par exemple, être approximé durant les périodes de silence. La définition de l'approche par soustraction spectrale est donnée par

$$\left|\hat{S}[\eta, l]\right| = \left|X[\eta, l]\right| - D_\mu \tag{2.11}$$

où D_μ correspond à l'espérance du bruit $D[\eta, l]$ lors des périodes de silence, soit $D_\mu = E\big[D[\eta, l]\big]$.

Un des gros désavantages de la soustraction spectrale telle que présentée est que l'estimation du bruit induit un nouveau bruit qui contient une certaine musicalité, nommé "bruit musical" [5]. Deux améliorations sont proposées dans [5] pour tenter de réduire ce bruit : premièrement, l'estimation du bruit soustrait est amplifiée par un facteur α plus grand que l'unité ; deuxièmement, un seuil minimal est fixé pour éviter que la soustraction cause un résultat plus bas qu'un certain niveau. Ce seuil correspond à une fraction de l'estimé du bruit, représenté par β. Ainsi, l'équation (2.11) devient :

$$\left|\hat{S}[\eta, l]\right|^\lambda = \begin{cases} \left|X[\eta, l]\right|^\lambda - \alpha[l]\left|\hat{D}[\eta, l]\right|^\lambda & \text{si } \left|X[\eta, l]\right|^\lambda > (\alpha[l] + \beta)\left|\hat{D}[\eta, l]\right|^\lambda \\ \beta\left|\hat{D}[\eta, l]\right|^\lambda & \text{sinon.} \end{cases} \tag{2.12}$$

où λ correspond à la puissance et sera expliqué plus bas, tandis que $\alpha[l]$ est donné par l'équation linéaire suivante

$$\alpha[l] = \alpha_0[l] - (\mathrm{RSB}_l)/s, \quad \text{pour } -5 \leq \mathrm{RSB} \leq 20. \tag{2.13}$$

$\alpha_0[l]$ correspond à la valeur de $\alpha[l]$ désirée pour un RSB de 0 dB, RSB_l correspond au RSB estimé de la trame l et $1/s$ correspond à la pente permettant de réduire la valeur de α plus le RSB est grand. Selon les auteurs, cette réduction permet de réduire la distorsion causée par la soustraction.

De manière expérimentale, les auteurs proposent d'avoir un facteur α entre 3 et 6 pour un RSB de 0 dB et un facteur de seuil β entre 0.005 et 1. Pour la pente de l'équation de α, une valeur recommandée est de 20/3, ce qui permet d'avoir un seuil de 1 pour α avec 4.75 comme plafond. La valeur de s est donc dépendante des seuils et plafonds imposés pour α.

En plus des deux améliorations proposées ci-haut, une modification supplémentaire a été faite à l'équation (2.11) qui consiste à pouvoir appliquer la soustraction à des valeurs différentes de puissance du spectre, qui est représentée par λ. Les auteurs recommandent d'utiliser la puissance du spectre, soit $\lambda = 2$. Cette valeur fait aussi consensus dans les autres types de rehaussement sur le spectre du signal, comme il sera vu plus bas. Les valeurs pour α sont bonnes uniquement lorsque la puissance du spectre est utilisée pour

la soustraction. En cas d'utilisation du module du spectre (donc $\lambda = 1$), il est suggéré d'utiliser une valeur de α comprise entre 2 et 2.2.

2.3.2 Rehaussement par estimateur spectral EQMM

Il est possible d'obtenir une représentation spectrale de (2.1) en y appliquant une TFD, on obtient alors sous forme polaire :

$$R[\eta, l] * e^{-j\angle X[\eta,l]} = A[\eta, l] * e^{-j\alpha[\eta,l]} + D[\eta, l] \tag{2.14}$$

où $A[\eta, l] = \Big|S[\eta, l]\Big|$, $R[\eta, l] = \Big|X[\eta, l]\Big|$ et $\alpha[\eta, l] = \angle S[\eta, l]$ correspondent au module de la composante spectrale η de S et X respectivement de même qu'à la phase de S. Il est posé que les coefficients A puissent être modélisés statistiquement comme une variable aléatoire gaussienne [11], le problème d'estimation de l'EQMM peut être réduit vers une estimation de A à partir d'un ensemble infini d'observations discrètes. Dans notre cas, ces observations sont les signaux bruités. De plus, comme les composantes spectrales sont considérées statistiquement indépendantes l'une de l'autre, il est nécessaire de disposer seulement de $X[\eta, l]$ pour obtenir l'estimateur EQMM du module. Ainsi, l'estimateur EQMM $\hat{A}$ de A est

$$\hat{A}[\eta, l] = E\{A[\eta, l]|X[\eta, l]\} \tag{2.15}$$

$$= \frac{\int_0^\infty \int_0^{2\pi} a[\eta, l]\, p\big(X[\eta, l]\Big|a[\eta, l], \alpha[\eta, l]\big)\, p\big(a_k, \alpha[\eta, l]\big)\, d\alpha[\eta, l]\, da[\eta, l]}{\int_0^\infty \int_0^{2\pi} p\big(X[\eta, l]|a[\eta, l], \alpha[\eta, l]\big)\, p\big(a[\eta, l], \alpha[\eta, l]\big)\, d\alpha[\eta, l]\, da[\eta, l]} \tag{2.16}$$

où $a[\eta, l] = A[\eta, l]$, $E\{\cdot\}$ correspond à l'espérance statistique et $p(\cdot)$ à la fonction de densité de probabilité (FDP). Sous le modèle statistique gaussien, $p(X\Big|a[\eta, l], \alpha[\eta, l])$ et $p(a[\eta, l], \alpha[\eta, l])$ sont respectivement donnés par

$$p(X[\eta, l]|a[\eta, l], \alpha[\eta, l]) = \frac{1}{\pi\lambda_d[\eta, l]} \exp\left\{-\frac{1}{\lambda_d[\eta, l]}\Big|X[\eta, l] - a[\eta, l] \exp^{j\alpha[\eta,l]}\Big|^2\right\} \tag{2.17}$$

$$p(a[\eta, l], \alpha[\eta, l]) = \frac{a[\eta, l]}{\pi\lambda_s[\eta, l]} \exp\left\{-\frac{a[\eta, l]^2}{\lambda_s[\eta, l]}\right\} \tag{2.18}$$

où $\lambda_s[\eta, l] \triangleq E\{\Big|S[\eta, l]\Big|^2\}$ et $\lambda_d[\eta, l] \triangleq E\{\Big|D[\eta, l]\Big|^2\}$ sont les variances de la $\eta^{\text{ième}}$ composante spectrale de la parole et du bruit respectivement. Ensuite en substituant (2.17) et

(2.18) dans (2.16), on obtient (voir Appendix A de [11]) :

$$\hat{A}[\eta,l] = \Gamma(1.5)\frac{\sqrt{\upsilon[\eta,l]}}{\gamma[\eta,l]}\exp\left\{-\frac{\upsilon[\eta,l]}{2}\right\}\left[(1+\upsilon[\eta,l])I_o\left(\frac{\upsilon[\eta,l]}{2}\right)+\upsilon[\eta,l]I_1\left(\frac{\upsilon[\eta,l]}{2}\right)\right]R[\eta,l]. \tag{2.19}$$

où $\Gamma(\cdot)$ est la fonction gamma avec $\Gamma(1.5) = \sqrt{\pi}/2$. $I_0(\cdot)$ et $I_1(\cdot)$ sont respectivement les fonctions modifiées de Bessel d'ordre zéro et un, $\upsilon[\eta,l]$ est défini par

$$\upsilon[\eta,l] \triangleq \frac{\xi[\eta,l]}{1+\xi[\eta,l]}\gamma[\eta,l] \tag{2.20}$$

où $\xi[\eta,l]$ et $\gamma[\eta,l]$ sont respectivement définis par

$$\xi[\eta,l] \triangleq \frac{\lambda_s[\eta,l]}{\lambda_d[\eta,l]} \tag{2.21}$$

et

$$\lambda[\eta,l] \triangleq \frac{R[\eta,l]^2}{\lambda_d[\eta,l]}. \tag{2.22}$$

$\xi_[\eta,l]$ et $\lambda[\eta,l]$ sont interprétés comme étant les RSB *a priori* et *a posteriori* respectivement. On obtient ensuite le signal rehaussé à l'aide de (2.19) et de (2.8)-(2.10).

Simplification à fort RSB

En observant le comportement de $\hat{A}[\eta,l]$ à de fort RSB ($\xi[\eta,l] \gg 1$), il s'avère que $\xi[\eta,l] \gg 1$ implique $\upsilon[\eta,l] \gg 1$ avec une forte probabilité, si on considère que $\upsilon[\eta,l]$ a une distribution exponentielle, c.-à-d. $p(\upsilon[\eta,l]) = 1/\xi[\eta,l]\exp(-\upsilon[\eta,l]/\xi[\eta,l])$. Alors, quand $\xi[\eta,l] \gg 1$, la fonction hypergéométrique confluente de (2.19) peut être remplacée par l'approximation suivante :

$$M(-0.5;1;-\upsilon[\eta,l]) \cong \frac{\sqrt{\upsilon[\eta,l]}}{\Gamma(1.5)}, \upsilon[\eta,l] \gg 1 \tag{2.23}$$

Ce qui donne :

$$\begin{aligned}\hat{A}[\eta,l] &\cong \frac{\xi[\eta,l]}{1+\xi[\eta,l]}R[\eta,l] \qquad \text{Fort RSB}\\ &\triangleq A^w[\eta,l].\end{aligned} \tag{2.24}$$

Comme $S[\eta,l] = A[\eta,l]\exp(j\alpha[\eta,l])$ est estimé par $\hat{S}[\eta,l] = \hat{A}[\eta,l]\exp(j\theta[\eta,l])$ où $\exp(j\alpha[\eta,l])$ est l'exponentielle complexe du signal bruité, la $\eta^{\text{ième}}$ composante spectrale peut être esti-

mée en utilisant l'approximation de l'équation (2.24)

$$\begin{aligned}\hat{S}[\eta,l] &\cong \frac{\xi[\eta,l]}{1+\xi[\eta,l]}X[\eta,l] \qquad \text{Fort RSB} \\ &\triangleq S^{w}[\eta,l].\end{aligned} \tag{2.25}$$

où le w indique en fait que l'estimateur EQMM à fort RSB correspond à un estimateur de Wiener.

2.3.3 Rehaussement par soustraction spectrale dans le domaine des modulations du spectre

La majeure partie des techniques de rehaussement ayant été développées opère dans le domaine spectral. De plus en plus de recherches physiologiques et psychoacoustiques semblent indiquer que le système auditif humain représente la parole sous forme spectrale, mais également en considérant la modulation de ce spectre. Ces modulations sont obtenues en prenant les variations temporelles des composantes spectrales d'un signal [38]. Les modulations du spectre peuvent aussi être vues comme la variable indépendante de la TFD d'une TFD [1].
Puisqu'il est relativement facile pour le système auditif de déceler un signal de parole dans un environnement bruité, il est tentant d'utiliser une représentation semblable à celle du système auditif afin de faire le rehaussement. Des approches de rehaussement opérant dans le domaine des modulations du spectre ont donc été proposées. Dans une de ces approches, un algorithme de rehaussement de la parole basé sur la soustraction des modulations du spectre est proposé [29].
En prenant le module du spectre du signal de parole, $\left|X[\eta,l]\right|$, tel que défini en (2.4), il est possible d'obtenir les modulations du spectre à partir d'une seconde TFD, ce qui donne

$$\mathcal{X}[\eta,l,m] = \sum_{\mathcal{K}=0}^{\mathcal{N}-1}\left|X[\eta,\mathcal{K}+l\mathcal{Z}]\right|v[\mathcal{K}]e^{-j\frac{2\pi m\mathcal{K}}{\mathcal{N}}} \tag{2.26}$$

où m, $\mathcal{N}$, $\mathcal{Z}$ et $v[\cdot]$ sont respectivement l'index de fréquence des modulations, la taille des trames de cette seconde TFD, l'index de décalage des trames modulatoires et la fonction de fenêtrage.
La soustraction des modulations du spectre est semblable à celle présentée pour la soustraction spectrale à la section 2.3.1. Le spectre modulatoire $\mathcal{X}[\eta,l,m]$ est converti en forme

polaire permettant ainsi d'obtenir le module des modulations du spectre tel que

$$\mathcal{X}[\eta, l, m] = \Big|\mathcal{X}[\eta, l, m]\Big| e^{j\angle\mathcal{X}[\eta,l,m]}. \tag{2.27}$$

Pour obtenir le signal de parole rehaussé, le module du spectre $\Big|\mathcal{X}[\eta, l, m]\Big|$ sera remplacé par son estimé rehaussé $\Big|\hat{S}[\eta, l, m]\Big|$ en utilisant la règle de soustraction présentée par [5]. Le spectre rehaussé sera obtenu par

$$\Big|\hat{S}[\eta, l, m]\Big| = \begin{cases} \left(\Big|\mathcal{X}[\eta, l, m]\Big|^{\gamma} - \rho\Big|\hat{\mathcal{D}}[\eta, l, m]\Big|^{\gamma}\right)^{\frac{1}{\gamma}}, & \text{si } \Big|\mathcal{X}[\eta, l, m]\Big|^{\gamma} - \rho\Big|\hat{\mathcal{D}}[\eta, l, m]\Big|^{\gamma} \geq \beta\Big|\hat{\mathcal{D}}[\eta, l, m]\Big|^{\gamma} \\ \left(\beta\Big|\hat{\mathcal{D}}[\eta, l, m]\Big|^{\gamma}\right)^{\frac{1}{\gamma}}, & \text{sinon} \end{cases} \tag{2.28}$$

L'estimation du module du spectre des modulations du bruit $\Big|\hat{\mathcal{D}}[\eta, l, m]\Big|$ est obtenue à partir des décisions prises par un algorithme de détection d'activité vocale (DAV), appliqué dans le domaine des modulations du spectre. Le DAV va classifier chaque trame et leur assigner une valeur de 1 en présence de bruit et 0 en cas de silence, en utilisant la règle suivante

$$\Phi[\eta, l] = \begin{cases} 1, & \text{si } \phi[\eta, l] \geq \theta \\ 0, & \text{sinon} \end{cases} \tag{2.29}$$

où $\phi[\eta, l]$ correspond au RSB d'un segment des modulations du spectre défini par

$$\phi[\eta, l] = 10\log_{10}\left(\frac{\sum_{m}\Big|\mathcal{X}[\eta, l, m]\Big|^2}{\sum_{m}\Big|\hat{\mathcal{D}}[\eta - 1, l, m]\Big|^2}\right) \tag{2.30}$$

et où θ correspond à un seuil de présence de parole déterminé de manière empirique. L'estimation du bruit est renouvelée selon l'équation de moyennage

$$\Big|\hat{\mathcal{D}}[\eta, l, m]\Big|^{\gamma} = \lambda\Big|\hat{\mathcal{D}}[\eta - 1, l, m]\Big|^{\gamma} + (1 - \lambda)\Big|\mathcal{X}[\eta, l, m]\Big|^{\gamma} \tag{2.31}$$

où λ est un facteur d'oubli défini de façon empirique selon la stationnarité du bruit. Le module rehaussé sera combiné avec la phase non modifiée pour donner l'estimation rehaussée du spectre modulatoire

$$\mathcal{Y}[\eta, l, m] = \Big|\hat{S}[\eta, l, m]\Big| e^{j\angle\mathcal{X}[\eta,l,m]}. \tag{2.32}$$

Pour obtenir le module du spectre rehaussé, la partie synthèse du modèle AMS est modifiée pour passer du domaine des modulations du spectre au domaine spectral. L'équation de synthèse est donnée par

$$\left|\hat{S}[\eta,l]\right| = \sum_{\mathcal{L}} \left\{ v[l-\mathcal{L}\mathcal{Z}] \sum_{m=0}^{\mathcal{N}-1} \mathcal{Y}[\eta,l,m] e^{j\frac{2\pi(l-\mathcal{L}\mathcal{Z})m}{\mathcal{N}}} \right\}. \tag{2.33}$$

Une fois le module du spectre obtenu, il est utilisé pour obtenir le signal temporel rehaussé à partir de l'équation (2.9).

2.3.4 Rehaussement par estimateur EQMM des modulations du spectre

Il a été montré dans le domaine spectral que le rehaussement par estimateur EQMM offre de meilleures performances que celui obtenu par la soustraction spectrale. Il a donc été proposé de développer un estimateur EQMM opérant dans le domaine des modulations du spectre plutôt que dans le domaine spectral [28]. De la même façon que pour le domaine spectral, le rehaussement par estimateur EQMM n'opère que sur le module du spectre modulatoire et ne tient pas compte de sa phase. De la même façon que pour l'estimateur obtenu par soustraction dans le domaine des modulations spectrales, une deuxième TFD est appliquée sur le module du spectre pour obtenir le signal donné en (2.27). L'approche EQMM consiste à déterminer une fonction de gain $\mathcal{G}[\eta,l,m]$ permettant de minimiser la différence entre le module du spectre modulatoire de la parole et son estimation. L'estimateur EQMM ainsi obtenu est :

$$\left|\hat{S}[\eta,l,m]\right| = \mathcal{G}[\eta,l,m]\left|\mathcal{X}[\eta,l,m]\right| \tag{2.34}$$

où $\left|\hat{S}[\eta,l,m]\right|$ correspond au module rehaussé et $\mathcal{G}[\eta,l,m]$ est la fonction de gain

$$\mathcal{G}[\eta,l,m] = \frac{\sqrt{\pi}}{2}\frac{\sqrt{v[\eta,l,m]}}{\gamma[\eta,l,m]}\Lambda[v[\eta,l,m]] \tag{2.35}$$

où $\Lambda[\cdot]$ correspond à

$$\Lambda[\theta] = e^{-\frac{\theta}{2}}\left[(1+\theta)I_0\left(\frac{\theta}{\theta}\right) + \theta I_1\left(\frac{\theta}{\theta}\right)\right] \tag{2.36}$$

où $I_0(\cdot)$ et $I_1(\cdot)$ correspondent aux fonctions modifiées de Bessel d'ordre zéro et un respectivement. Ensuite, $v[\eta, l, m]$ est donné par

$$v[\eta, l, m] \triangleq \frac{\epsilon[\eta, l, m]}{1 + \epsilon[\eta, l, m]} \gamma[\eta, l, m]. \tag{2.37}$$

$\epsilon[\eta, l, m]$ et $\gamma[\eta, l, m]$ sont respectivement le RSB *a priori* et *a posteriori* et sont définis par

$$\epsilon[\eta, l, m] \triangleq \frac{E\left[\left|\mathcal{S}[\eta, l, m]\right|^2\right]}{\left|\mathcal{D}[\eta, l, m]\right|^2} \tag{2.38}$$

et

$$\gamma[\eta, l, m] \triangleq \frac{\left|\mathcal{X}[\eta, l, m]\right|^2}{\left|\mathcal{D}[\eta, l, m]\right|^2}. \tag{2.39}$$

Comme il n'est pas possible de récupérer le RSB *a priori* et *a posteriori*, $\epsilon[\eta, l, m]$ et $\gamma[\eta, l, m]$ sont estimés par la même technique de décision dirigée pour le domaine spectral décrit en [11], mais modifiée pour être appliquée dans le domaine des modulations de spectre. L'estimé de $\epsilon[\eta, l, m]$ est donné par

$$\hat{\epsilon}[\eta, l, m] = \alpha \frac{\left|\hat{\mathcal{S}}[\eta, l, m-1]\right|^2}{\hat{\lambda}[\eta, l, m-1]} + (1-\alpha)\max[\hat{\gamma}[\eta, l, m] - 1, 0] \tag{2.40}$$

où α est un paramètre empirique permettant de choisir le compromis entre la réduction de bruit et la distorsion transitoire du signal et où $\hat{\lambda}[\eta, l, m]$ est l'espérance de l'énergie du spectre modulatoire de l'estimation du bruit. Ensuite, le RSB *a posteriori* est estimé par

$$\hat{\gamma}[\eta, l, m] = \frac{\left|\mathcal{X}[\eta, l, m]\right|^2}{\hat{\lambda}[\eta, l, m]}. \tag{2.41}$$

La limitation de la valeur minimale du RSB *a priori* ajoute un résidu de bruit après rehaussement, ce qui cause une distorsion non négligeable [11] [9]. Pour pallier à ce problème, un seuil est déterminé pour éviter d'obtenir une valeur causant ces distorsions. Le RSB *a priori* est donc donné par

$$\hat{\gamma}[\eta, l, m] = \max[\hat{\gamma}[\eta, l, m], \gamma_{\min}] \tag{2.42}$$

où $\gamma_{\min}$ corresponds au seuil minimal à atteindre. Une valeur empirique de -25 dB est utilisée pour ce seuil, sans précision par les auteurs de la démarche ayant permis d'atteindre cette valeur. Finalement, le module du spectre modulatoire rehaussé sera utilisé dans l'équation (2.33) qui donnera le spectre rehaussé utilisé dans l'équation (2.9) permettant d'obtenir le signal de parole rehaussé.

2.3.5 Rehaussement avec la technique de Fusion

Les travaux en [28] proposent une méthode de rehaussement combinant le rehaussement fait dans le domaine spectral avec celui fait dans le domaine des modulations du spectre. Cette combinaison se fait avec une fonction de poids permettant d'utiliser un domaine plus que l'autre selon le RSB *a posteriori* de la trame à rehausser. Cette combinaison se fait dans le domaine spectral, ce qui veut dire que le rehaussement fait dans le domaine des modulations du spectre doit être converti vers le domaine spectral avant de pouvoir être utilisé. La fonction de poids est donnée par

$$\left|\hat{S}[\eta,l]\right| = \left(\Psi[\omega[l]]\left|\hat{S}_{\text{spectral}}[\eta,l]\right|^{\lambda} + (1-\Psi[\omega[l]])\left|\hat{S}_{\text{modulatoire}}[\eta,l]\right|^{\lambda}\right)^{\frac{1}{\lambda}} \tag{2.43}$$

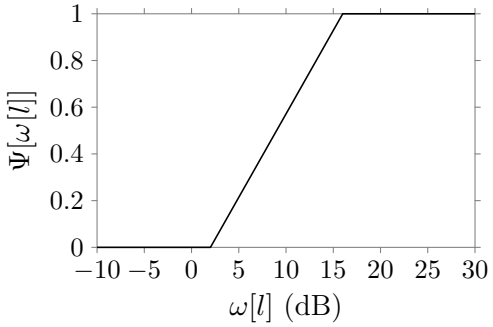

Figure 2.1 Fonction de poids, $\Psi[\omega[l]]$

où le module du spectre rehaussé $\left|\hat{S}[\eta,l]\right|$ est obtenu par une combinaison du module du spectre rehaussé $\left|\hat{S}_{\text{spectral}}[\eta,l]\right|$ et du module des modulations spectrales $\left|\hat{S}_{\text{modulatoire}}[\eta,l]\right|$. $\Psi[\omega[l]]$ correspond au facteur de poids donné pour la combinaison et $\omega[l]$ correspond au

RSB de la trame l. $\Psi[\omega[l]]$ est donné par

$$\Psi\big[\omega[l]\big] = \begin{cases} 0 & \text{si } g\big[\omega[l]\big] \leq 2 \\ \frac{g\left[\omega[l]\right]-2}{14} & \text{si } 2 < g\big[\omega[l]\big] < 16 \\ 1 & \text{si } g\big[\omega[l]\big] \geq 16 \end{cases} \tag{2.44}$$

où $g\big[\omega[l]\big] = 10\log_{10}\big[\omega[l]\big]$ et où $\omega[l]$ correspond au RSB après rehaussement. La fonction de poids est illustrée à la fig. 2.1. Les trames ayant un RSB plus bas que 2 dB seront rehaussés uniquement dans le domaine des modulations du spectre, celles ayant un RSB plus grand que 16 dB seront rehaussées uniquement dans le domaine spectral tandis que les trames entre ces deux limites seront rehaussées par une combinaison des deux domaines.

La technique de rehaussement utilisée pour le domaine spectral pour obtenir $\hat{S}_{\text{spectral}}[\eta, l]$ dans [28] est l'algorithme de rehaussement par estimateur EQMM du module du spectre tandis que l'algorithme pour le rehaussement du module des modulations du spectre $\hat{S}_{\text{modulatoire}}[\eta, l]$ est celui du rehaussement par soustraction spectrale.

Cette technique de combinaison offre une amélioration des performances étant donné que les résultats avec le rehaussement spectral sont meilleurs que ceux avec le rehaussement des modulations du spectre à fort RSB tandis qu'à faible RSB, le rehaussement des modulations du spectre performe mieux que le rehaussement spectral. Le tableau 2.1 ci-dessous contient les résultats obtenus par les auteurs de [29] pour la comparaison de leur algorithme avec le rehaussement spectral et des modulations du spectre. Les performances sont mesurées à partir du score PESQ, présenté en section 4.1.1.

Tableau 2.1 Score PESQ avec un bruit additif blanc gaussien (BABG).

SNR (dB)	*MMSE*	*ModSpecSub*	*SpecSub*	*Fusion*
0	2.00	2.22	1.76	2.25
5	2.10	2.49	2.18	2.63
10	2.62	2.76	2.62	2.79
15	2.97	3.02	3.01	3.05

Le terme *MMSE* est la traduction anglaise de EQMM (*Minimum Mean Square Error*). Il correspond à la technique présentée en section 2.3.2. Le terme *ModSpecSub* correspond à la technique présentée en section 2.3.3. La technique *SpecSub* est décrite à la section 2.3.1 et *Fusion* est la technique de rehaussement utilisant la fonction de combinaison présentée ci-dessus.

2.4 Estimation du bruit par estimation des statistiques minimales du bruit

Pour que les algorithmes de rehaussement puissent fonctionner correctement, il est nécessaire de récupérer une estimation de la densité spectrale de puissance du bruit. Il s'agit d'une partie importante étant donné que la performance de l'algorithme de rehaussement dépend de la performance de l'estimation de la densité spectrale de puissance (DSP) du bruit, surtout en présence de bruit non stationnaire. En effet, si l'estimation est trop faible, du bruit résiduel articiel sera toujours présent. Si l'estimation est trop forte, il y aura alors distorsion de la parole avec une perte d'intelligibilité.

Les premières techniques d'estimation faisaient appel à un détecteur d'activité vocale (DAV) et tentaient d'estimer le bruit lors de périodes de silence. Le problème principal de ces estimateurs est qu'il est difficile de les configurer de manière générale et que leur performance est faible lors de faible RSB [15] [22]. La technique développée par [23] propose d'estimer la DSP du bruit à partir d'une méthode optimale de lissage de la DSP et du principe de statistiques minimales. Le fait d'utiliser un algorithme de statistiques minimales permet d'éviter d'avoir un seuil minimal fixe et permet plutôt de disposer d'un seuil variable dans le temps et indépendant pour chaque composante spectrale. Il a d'ailleurs été confirmé que cette approche offre une bonne performance dans le cas de bruit non stationnaire [25]. Cette approche est présentée plus en détail dans ce qui suit.

L'algorithme de statistiques minimales repose sur deux observations: 1 - la parole et le bruit sont statistiquement indépendants et 2 - la puissance de signal de parole bruité s'ajuste fréquemment au niveau de puissance du bruit lui-même. La première observation correspond aux observations faites pour la plupart des algorithmes de rehaussement de la parole tandis que la seconde observation tient compte du fait qu'il y a beaucoup de silence dans un signal de parole. Ainsi, il est possible d'estimer la DSP du signal de bruit en tendant vers la DSP minimale du signal de parole bruité.

2.4.1 Principes d'estimation des statistiques minimales du bruit

Cette section présente l'équation permettant de trouver l'estimation des statistiques minimales de la DSP du bruit. Elle comprend un facteur de correction $\rho[\eta, l]$ différent pour chaque composante spectrale. De plus, le facteur optimal à avoir pour obtenir l'estimation de la DSP minimale du bruit sera trouvé en localisant le zéro de la dérivée de l'équation par rapport au facteur de correction $\rho[\eta, l]$.

Comme les coefficients de la TF peuvent être considérés indépendants et peuvent être estimés par une variable aléatoire gaussienne à moyenne nulle [8], la FDP du module de la puissance spectrale du signal de parole bruitée $\left|X[\eta,l]^2\right|$ est donnée par:

$$f_{\left|X[\eta,l]^2\right|}(x) = \frac{1}{\lambda_s[\eta,l]+\lambda_d[\eta,l]} e^{\lambda_s[\eta,l]+\lambda_d[\eta,l]} \tag{2.45}$$

où $\lambda_s[\eta,l] \triangleq E\{\left|S[\eta,l]\right|^2\}$ et $\lambda_d[\eta,l] \triangleq E\{\left|D[\eta,l]\right|^2\}$ sont les variances de la $\eta^{\text{ième}}$ composante spectrale de la parole et du bruit respectivement. Lors de la présence de silence, la variance du signal de parole étant nulle, la moyenne et la variance de $\left|X[\eta,l]^2\right|$ sont $\lambda_d[\eta,l]$ et $\lambda_d[\eta,l]^2$ respectivement. À partir de cette observation, l'algorithme de statistiques minimales va tenter de trouver la valeur minimale de la DSP, qui devrait correspondre à la DSP du bruit. Ainsi, lors de silence entre les phrases et mots, là où l'énergie du signal de parole est nulle, l'algorithme va trouver l'énergie du bruit. Ainsi, pour estimer la DSP du bruit, la puissance spectrale est moyennée de manière récursive, tel que

$$P[\eta,l] = \rho[\eta,l]P[\eta-1,l] + (1-\rho[\eta,l])\left|X[\eta,l]^2\right| \tag{2.46}$$

où $P[\eta,l]$ correspond à l'estimation de la puissance du bruit et $\rho[\eta,l]$ correspond au facteur de correction permettant de modifier le niveau d'énergie de chaque composante spectrale.

Comme on cherche à obtenir le plus possible la DSP du bruit, on cherche donc à minimiser l'erreur quadratique moyenne conditionnelle

$$E\Big\{(P[\eta,l]-\lambda_d[\eta,l])^2\Big|P[\eta-1,l]\Big\} \tag{2.47}$$

d'une itération à une autre. En substituant $P[\eta-1,l]$ de (2.46) dans (2.47) et en utilisant $E\{\left|X[\eta,l]^2\right|\} = \lambda_d[\eta,l]$ ainsi que $E\{\left|X[\eta,l]^4\right|\} = \lambda_d^2[\eta,l]$, l'erreur quadratique moyenne devient

$$E\Big\{(P[\eta,l]-\lambda_d[\eta,l])^2\Big|P[\eta-1,l]\Big\} = \rho[\eta,l]^2(P(\eta-1,l)-\lambda_d[\eta,l]^2)^2 + \lambda_d[\eta,l]^4(1-\rho[\eta,l])^2. \tag{2.48}$$

En posant la dérivée de (2.48) par rapport à $\rho[\eta,l]$ à zéro, on obtient

$$\rho_{\text{opt}}[\eta,l] = \frac{1}{1+(P[\eta-1,l]/\lambda_d[\eta,l]-1)^2} \tag{2.49}$$

où le terme $P(\eta-1,l)/\lambda_d[\eta,l]=\gamma[\eta,l]$ correspond au RSB *a posteriori* [24]

$$\gamma[\eta,l]=\frac{\left|Y[\eta-1,l]\right|^2}{\lambda_d[\lambda,l]}. \tag{2.50}$$

Finalement, comme la dérivée seconde de (2.49) par rapport à $\rho[\eta,l]$ sera non-nulle, ceci indique donc un minimum.

2.4.2 Optimisation du facteur de correction en fonction de l'erreur d'estimation

Cette section porte sur la problématique de devoir utiliser une estimation de la DSP réelle du bruit dans le calcul du facteur de correction $\rho[\eta,l]$. L'utilisation d'une estimation à pour conséquence d'induire une erreur dans le calcul de ce facteur, réduisant la performance de l'algorithme. Une méthode est présentée pour réduire l'impact de cette erreur dans le calcul du facteur de correction.

Lors d'une implémentation en temps réel, il est impossible de trouver la véritable DSP du bruit $\lambda_n[\eta,l]$. Il faut donc la remplacer par l'estimé le plus proche, soit la DSP de la trame précédente $\lambda_n[\eta-1,l]$, puis limiter la valeur maximale du facteur de correction à une valeur $\rho_{\max}$ (par exemple, $\rho_{\max}=0.96$) pour éviter l'impasse causée par $\gamma[\eta,l]=1$ ($\gamma[\eta,l]=1$ donne $\rho=1$, empêchant la puissance estimée d'être mise à jour dans (2.46)). Généralement, l'évolution temporelle de $\lambda_n[\eta,l]$ sera toujours en retard par rapport à la DSP réelle. Ceci cause donc une estimation de la DSP du bruit qui sera fort ou trop faible, amenant à un facteur de lissage trop petit ou trop gros. Lorsque le paramètre de lissage approche de l'unité, il lui sera impossible de répondre rapidement à des changements soudains de la DSP.

Ainsi, à cause de cette possibilité, il faut suivre l'erreur présente dans l'estimateur $P[\eta,l]$ à court terme. Si une erreur est détectée, il faut modifier le facteur de correction pour s'assurer que l'algorithme puisse continuer à fonctionner correctement. Une manière de surveiller cette erreur est de comparer $P[\eta,l]$ à une référence, soit la moyenne combinée de chaque composante spectrale de la DSP. L'algorithme de surveillance va comparer la moyenne à court terme de la trame précédente $1/L\sum_{l=0}^{L-1}P[\eta-1,l]$ par rapport à la moyenne $1/L\left|Y[\eta,l]\right|^2$. Une déviation trop grande de l'estimation à court terme sera donc

détectée et sera corrigée en modifiant le facteur de correction avec le paramètre suivant:

$$\tilde{\rho}_c[\eta] = \frac{1}{1 + \left(\sum_{l=0}^{L-1} P[\eta-1, l] / \sum_{l=0}^{L-1} \left|Y[\eta, l]\right|^2 - 1\right)^2}. \tag{2.51}$$

De plus, le facteur de correction sera limité à des valeurs plus grandes que 0.7 et sera moyenné temporellement

$$\rho[\eta] = 0.7\rho[\eta-1, l] + 0.3\max(\tilde{\rho}_c[\eta], 0.7). \tag{2.52}$$

Les valeurs utilisées en (2.52) ont été trouvées de manière empirique et semblent être robustes. À partir du facteur de correction défini en (2.49), le paramètre de correction a_c et le paramètre du maximum $\rho_{\max}$ sont ajoutés pour donner l'équation finale

$$\hat{\alpha}[\eta, l] = \frac{\alpha_{\max}\alpha_c[\eta]}{1 + (P(\eta-1, l)/\lambda_n[\eta-1, l] - 1)^2}. \tag{2.53}$$

Le problème du facteur de correction défini en (2.53) est qu'il est sous-optimal par rapport à la version présentée en (2.49). Pour un bruit stationnaire, la déviation par rapport à ρ_{opt} est de 5 % tandis qu'elle est de 10 % pour les bruits non stationnaires, comme les bruits de ville.

Pour tenter d'augmenter les performances de l'estimateur de bruit lorsque celui-ci est fort et non-stationnaire, il est nécessaire de définir un seuil minimal pour le paramètre de lissage, $\rho_{\min}$, avec un maximum de 0.3. Cette limite peut causer une dégradation des performances de la présence d'un signal à fort RSB. Comme $\rho_{\min}$ freine le temps de montée et de descente de $P[\eta, l]$, une autre limite minimale est instaurée en fonction du RSB moyen du signal de parole bruité. Pour éviter d'atténuer les consonnes faibles à la fin d'un mot, il faut s'assurer que $P[\eta, l]$ puisse diminuer de sa valeur maximale jusqu'au niveau du bruit en environ 64 ms. Ce qui donne comme définition de $\rho_{\min}$

$$\rho_{\min} = \min(0.3, \text{RSB}^{-\frac{R}{0.0064 f_s}}) \tag{2.54}$$

où R correspond à la moitié du nombre d'échantillons dans une trame ($N = 2R$).

2.4.3 Estimateur de bruit non biaisé de statistique minimale

Cette section présente une modification de l'estimateur (2.50) pour le rendre non biaisé. Comme la valeur minimale d'un ensemble de variables aléatoires est plus petite que la moyenne de cet ensemble, l'estimateur minimum du bruit est nécessairement biaisé. En

trouvant le biais et la variance de l'estimateur, il est possible de développer un algorithme pour la compensation de ce biais en présence de bruit non stationnaire. Le développement ci-dessous suppose que la séquence successive de $P[\eta, l]$ est corrélée et qu'une solution approximative sera trouvée. La preuve permettant d'atteindre l'estimateur non biaisé n'est pas présentée ci-dessous. Le lecteur est invité à consulter la démarche de l'auteur dans ses travaux [23].

L'estimateur non biaisé de la DSP du bruit $\lambda_n[\eta, l]$ est donné par

$$\hat{\lambda}_n[\eta, l] = B_{\min}[\eta, l] P_{\min}[\eta, l] \tag{2.55}$$

où $B_{\min}$ est le compensateur de biais et $P_{\min}[\eta, l]$ correspond à la DSP minimale des D dernières trames. Le compensateur $B_{\min}[\eta, l]$ est défini par

$$B_{\min}[\eta, l] \approx 1 + (D-1)\frac{2}{\tilde{Q}_{\text{eq}}[\eta, l]} \tag{2.56}$$

où $Q_{\text{eq}}[\eta, l]$ correspond au "degré de liberté équivalent" du compensateur de biais. Il correspond à l'inverse de la variance normalisée. Il est défini par

$$\frac{1}{Q_{\text{eq}}[\eta, l]} \approx \frac{\hat{\text{var}}\{P[\eta, l]\}}{2\lambda_n^2[\eta - 1, l]}. \tag{2.57}$$

La variance estimée est trouvée à partir d'une approximation des moments de premier et second ordre de l'espérance de $P[\eta, l]$

$$\hat{\text{var}}\{P[\eta, l]\} = \bar{P}^2[\eta, l] - \bar{P}^2[\eta, l]. \tag{2.58}$$

où $\bar{P}[\eta, l]$ et $\bar{P}^2[\eta, l]$ sont défini par

$$\bar{P}[\eta, l] = \beta[\eta, l]\bar{P}[\eta - 1, l] + (1 - \beta[\eta, l])P[\eta, l] \tag{2.59}$$

$$\bar{P}^2[\eta, l] = \beta[\eta, l]\bar{P}^2[\eta - 1, l] + (1 - \beta[\eta, l])P^2[\eta, l]. \tag{2.60}$$

Dans l'estimation de la variance, $\beta[\eta, l]$ est défini par

$$\beta[\eta, l] = \min(\rho^2, 0.8). \tag{2.61}$$

Finalement, le calcul de l'estimé de la variance d'une trame de la DSP (2.55) utilise une version à l'échelle de $Q_{\text{it}}[\eta, l]$, $\tilde{Q}_{\text{it}}[\eta, l]$, qui est défini par

$$\tilde{Q}_{\text{eq}}[\eta, l] = \frac{Q_{eq}[\eta, l] - 2M(D)}{1 - M(D)} \tag{2.62}$$

où $M[\cdot]$ est une fonction de D définie dans les travaux de l'auteur (voir *Appendix B* de [23]).

CHAPITRE 3

Méthodologie

Ce projet de recherche consiste à développer un algorithme de rehaussement de la parole combinant les domaines spectral et de modulations du spectre. Il est basé sur la fonction de combinaison présentée à l'équation (2.43), où nous utiliserons un estimateur EQMM dans le domaine des modulations spectrales pour $\hat{S}_{\text{modulatoire}}[\eta, l]$ au lieu de l'estimateur par soustraction spectrale utilisée dans [28]. Cet algorithme sera comparé aux différentes techniques de rehaussement présentées dans le chapitre précédent afin de déterminer si notre approche offre de meilleures performances que ces dernières. De plus, tous les algorithmes, à l'exception de l'algorithme présenté en 2.3.1, seront modifiés pour fonctionner avec une fréquence d'échantillonnage de 16 kHz. Toutes les techniques utilisent le même estimateur de bruit développé en [23], mais modifié pour fonctionner avec une fréquence d'échantillonnage de 16 kHz ainsi que dans le domaine des modulations du spectre. Cela permet de s'assurer que l'on compare uniquement les performances des algorithmes de rehaussement.

L'algorithme de rehaussement proposé opère dans les deux domaines présentés. Un estimateur EQMM sera utilisé pour chaque domaine compte tenu de la performance de cet estimateur dans chaque domaine [11] [28] respectif. La fonction de combinaison présentée dans [29] sous le nom de Fusion sera modifié pour tenir compte des changements de cadence et d'algorithme de rehaussement dans le domaine des modulations du spectre.

Les algorithmes de rehaussement fusionnant l'estimateur EQMM de l'amplitude spectrale (EQMM AS), présenté en section 2.3.2, et l'estimateur EQMM du module des modulations spectrales (EQMM MMS), présenté en section 2.3.4 avec la nouvelle fonction de combinaison représente le nouvel algorithme de rehaussement présenté dans ces travaux. Cet ensemble est nommé EQMM double à partir de maintenant.

La suite du chapitre présentera les éléments suivants: la conception de la banque de segments bruités pour la mesure de performance des algorithmes de rehaussement, la modification des différents algorithmes pour le changement de fréquence d'échantillonnage de 8 kHz vers 16 kHz, puis la modification de la fonction de Fusion présentée en [29].

3.1 Composition de la banque de segments bruités

La banque de segments de parole bruités utilise la base de données de parole du *Telecommunication and signal processing laboratory* du Département de génie électrique et de génie informatique de l'Université McGill [19]. Il s'agit d'une base de données contenant 1440 courtes phrases parlées par 24 personnes différentes (12 femmes et 12 hommes). Ces segments sont ensuite bruités, à l'aide de la base de données de bruit NOISEX-92 [36], afin de représenter différents RSB.

3.1.1 Base de données de segment de paroles

La base de données utilise la liste de phrases venant de l'université Harvard [10] comme source. Les phrases sont séparées en 72 listes, chacune contenant 10 phrases. Chaque personne lit 6 listes, soit 60 phrases. Ces phrases sont en anglais et la majorité des personnes parle l'anglais canadien. Les autres locuteurs parlent un autre dialecte ou n'ont pas l'anglais comme langue maternelle.
Les enregistrements ont eu lieu dans une chambre anéchoïque avec le microphone légèrement décalé pour qu'il ne soit pas directement en face du locuteur, évitant d'enregistrer les bruits associés à la respiration.
La banque de données a été échantillonnée à 48 kHz. Ensuite, elle a été sous-échantillonnée à 16 kHz et 8 kHz. La version à 16 kHz a été sous-échantillonnée avec un filtre passe-bas RIF d'ordre 168. Ce filtre provient de la librairie d'outils logiciels du *Internation Telecommunication Union - Telecommunication* ITU-T [18]. La version à 8 kHz est sous-échantillonnée à partir des données à 16 kHz et d'un filtre de type *modified IRS-send* [19].

3.1.2 Banque de bruits

Tel qu'indiqué, le nom de la base de données utilisée pour les signaux de bruits est NOISEX-92 [36]. Cette base de données contient différents bruits avec des statistiques stationnaires et non-stationnaires. Les types de bruits inclus dans la base de données sont:

- Bruits de conversations
- Bruits d'usine
- Bruits HF radio, rose et blanc
- Divers bruits militaires (bruits d'avion de chasse, de char d'assaut et de mitraillette)
- Bruits de voiture

Ces bruits ont une fréquence d'échantillonnage de 18 kHz. Ils ont été décimés à 8 et 16 kHz pour être correctement mélangés avec les segments de paroles.

3.1.3 Mixage des segments de parole avec les segments de bruit

Pour chaque fréquence d'échantillonnage, la parole est mixée avec du bruit afin d'obtenir des RSB de 0, 5, 10 et 15 dB. Ces quatre valeurs de RSB permettent de tester une plage réaliste de RSB présent dans un environnement non contrôlé. Il s'agit aussi de la plage utilisée pour calculer la performance d'autres algorithmes de rehaussement de la parole. Pour le mixage, le niveau du signal de parole est déterminé à partir de la spécification P.56 de l'ITU-T [33] tandis que le niveau de bruit est déterminé par son énergie.

3.2 Changement de cadence de 8 kHz à 16 kHz

Avec l'amélioration des réseaux téléphoniques commutés (RTC) traditionnels et des réseaux cellulaires permettant d'utiliser des codecs à large bande, il est nécessaire de s'assurer que l'algorithme de rehaussement EQMM double proposé puisse fonctionner correctement à une fréquence d'échantillonnage de 16 kHz, en plus d'une fréquence d'échantillonnage de 8 kHz.

3.2.1 Déploiement des RTC et réseaux cellulaires à large bande

Les RTC, de par leur technologie, utilisent un codec audio à bande étroite, le G.711 [32], qui, initialement, avait une plage dynamique de 300 à 3400 Hz. En 2008, ce codec a été mis à jour pour qu'il devienne à bande large, soit une plage dynamique de 50 à 7000 Hz avec une fréquence d'échantillonnage de 16 kHz [34]. Il n'est pas utilisé sur les RTC analogiques, étant donné les contraintes venant de l'équipement ne supportant pas une telle fréquence d'échantillonnage. Par contre, ces réseaux analogiques sont remplacés de plus en plus par des réseaux numériques connectés à Internet. Ces réseaux ont la capacité d'accepter des codecs avec un débit de 16 kbit/s, permettant en même temps de profiter d'une fréquence d'échantillonnage de 16 kHz.
Les réseaux cellulaires permettaient initialement le transfert de parole échantillonnée à 8 kHz afin d'être compatible avec les RTC. L'arrivée récente de la technologie VoLTE (*Voice over LTE*) permet d'utiliser le codec large bande AMR-WB (*Adaptive Multi-Rate Wideband*) [13]. Ce codec utilise une fréquence d'échantillonnage de 16 kHz. Une fois les réseaux 3G retirés, les appels téléphoniques cellulaires utiliseront uniquement la VoLTE. En date du 13 novembre 2017, tous les opérateurs majeurs canadiens offrent la VoLTE sur leurs réseaux [3].

3.2.2 Modification des algorithmes de rehaussement

Tous les algorithmes utilisés dans ces travaux ont été développés initialement avec une fréquence d'échantillonnage de 8 kHz. Il est donc nécessaire de modifier leurs différents paramètres et de les valider pour s'assurer d'avoir un niveau de performance comparable une fois la fréquence d'échantillonnage à 16 kHz. En théorie, ces algorithmes sont indépendants de la fréquence d'échantillonnage. Mais comme la majeure partie de l'énergie de la parole se situe en dessous de 4 kHz, il est possible que l'augmentation de la fréquence d'échantillonnage réduise le RSB pour une même situation, étant donné le peu d'information de parole ajouté. Pour la calibration de ces paramètres, un bruit additif blanc gaussien (BABG) est ajouté aux segments de parole à 0, 5, 10 et 15 dB. La mesure de performance utilise le score PESQ, qui sera présenté à la section 4.1.1. Nous présentons dans ce qui suit la variation des paramètres des différents estimateurs utilisés dans ce mémoire pour une fréquence d'échantillonnage de 16 kHz.

Rehaussement par estimateur EQMM dans le domaine spectral

Les paramètres étudiés pour l'estimateur EQMM dans le domaine spectral sont: longueur de trame spectrale, facteur de lissage α et le seuil minimum $\gamma_{\min}$ pour le RSB *a priori*.

Longueur des trames spectrales

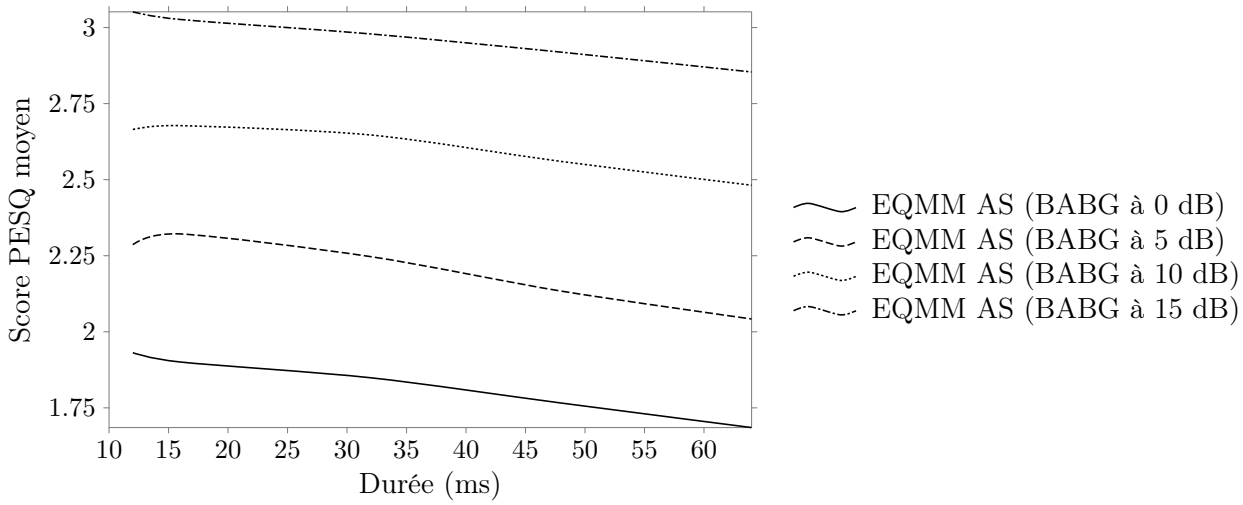

Figure 3.1 Effet de la durée d'une trame spectrale sur le score PESQ pour l'algorithme EQMM AS. Une durée de trame d'environ 15 ms permet d'avoir le meilleur score.

Sous 8 kHz, la longueur de trame doit se situer entre 20 et 30 ms pour avoir une performance de rehaussement optimale. Dans le cas de la figure 3.1, le score PESQ atteint son maximum à environ 15 ms, qui reste très proche quand même des valeurs recommandées. La valeur finale utilisée dans l'algorithme EQMM double pour la longueur de trame spectrale est de 16 ms, pour rester dans les puissances de 2. Ce choix d'obtenir de meilleures performances de calcul, car les différents algorithmes informatiques sont optimisés pour des vecteurs ayant une taille faisant partie des puissances de 2.

Facteur de lissage α

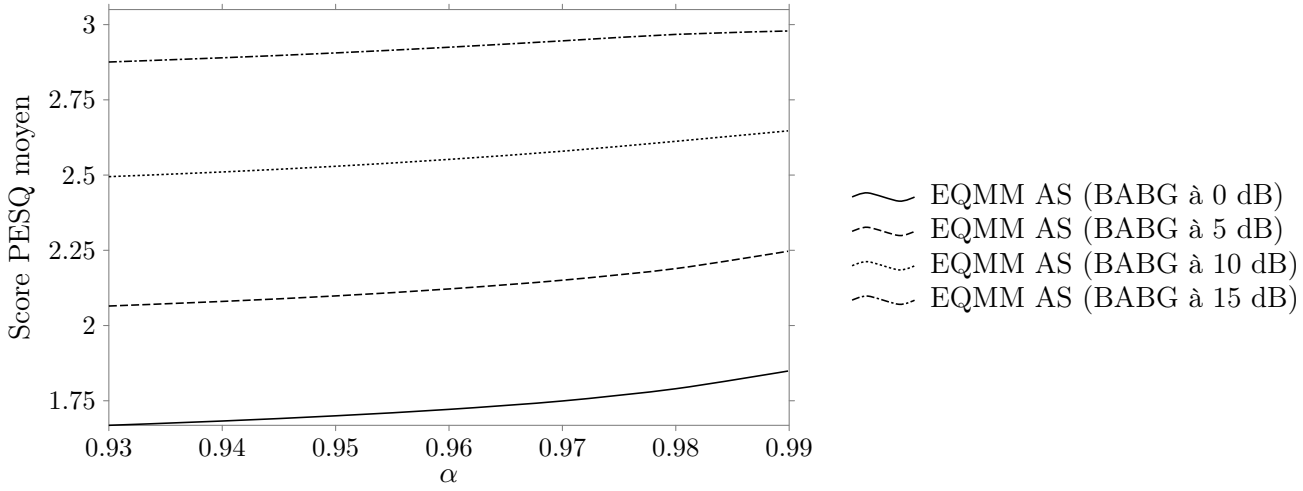

Figure 3.2 Effet du facteur de lissage α sur le score PESQ pour l'algorithme EQMM AS. Une valeur de 0.99 permet d'avoir le meilleur score PESQ.

Les travaux de [5] concluaient que la valeur optimale pour α était 0.99. Dans la figure 3.2, cette valeur reste donc identique pour une fréquence d'échantillonnage de 16 kHz pour tout RSB. Ce paramètre sert à choisir de manière empirique entre la réduction de bruit et la distorsion transitoire du signal. Il est donc normal que la fréquence d'échantillonnage n'ait pas d'influence sur ce paramètre.

Seuil minimum $\gamma_{\min}$ pour le RSB *a priori*

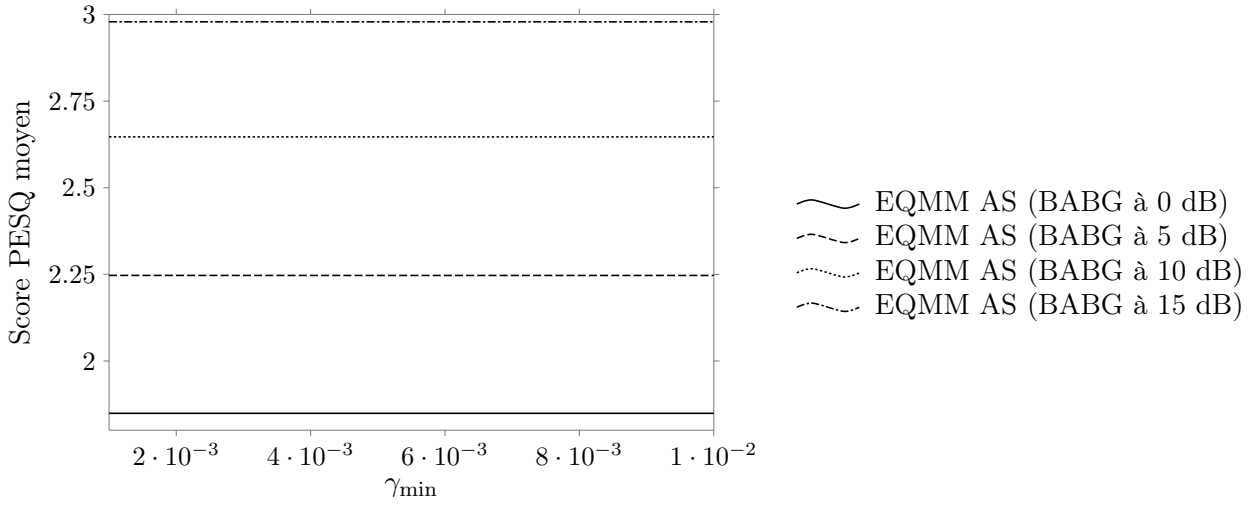

Figure 3.3 Effet du seuil minimum $\gamma_{\min}$ pour le RSB *a priori* sur le score PESQ pour l'algorithme EQMM AS. La variation du seuil minimum n'a aucune influence sur le score PESQ.

Le seuil minimum $\gamma_{\min}$, présenté à la figure 3.3 sert à éviter d'avoir une réduction du bruit trop agressif. Parmi les valeurs optimales à 8 kHz, allant de 0.001 à 0.01, la variation de ce paramètre ne change aucunement la performance de l'algorithme. L'algorithme EQMM AS n'atteint pas son cas limite en présence d'un BABG. Ce paramètre pourrait faire varier sa performance en présence d'un autre bruit.

Rehaussement par estimateur EQMM dans le domaine des modulations du spectre

Les paramètres à valider pour l'estimateur EQMM dans le domaine des modulations du spectre sont: longueur de trame spectrale, longueur de trame modulatoire, facteur de lissage α et le seuil minimum $\gamma_{\min}$ pour le RSB *a priori*.

Longueur des trames spectrales

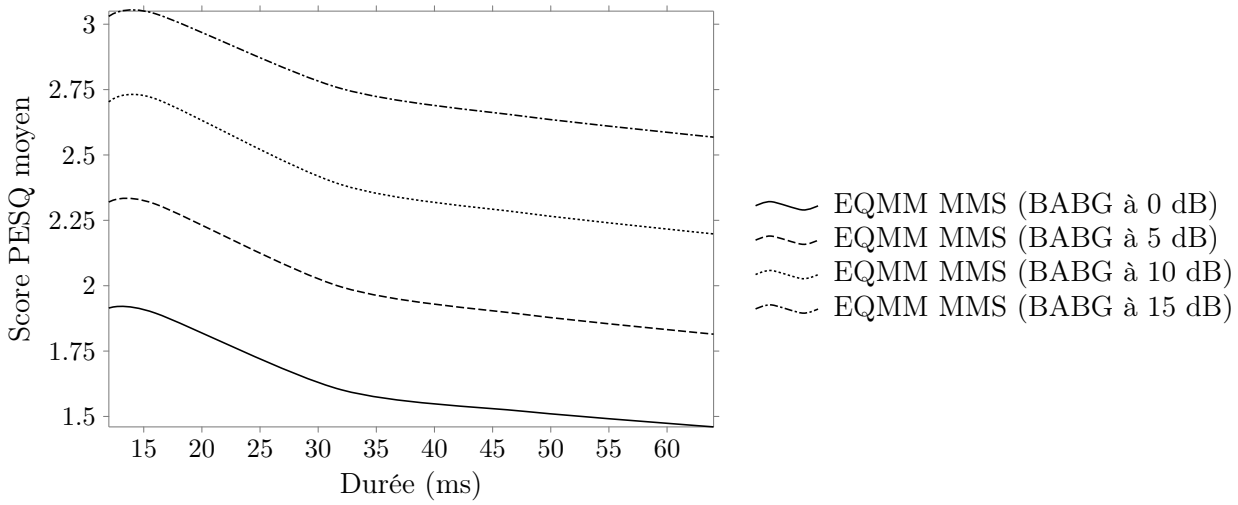

Figure 3.4 Effet de la durée d'une trame spectrale sur le score PESQ pour l'algorithme EQMM MMS. Une durée de trame de 15 ms permet d'avoir le meilleur score PESQ.

La figure 3.4 affiche une performance maximale autour de 15 ms, comme avec l'algorithme EQMM AS, mais avec une dégradation des performances plus accrues lorsque la durée augmente. Ce résultat permet donc de continuer à utiliser des trames spectrales de 16 ms.

Longueur des trames modulatoires

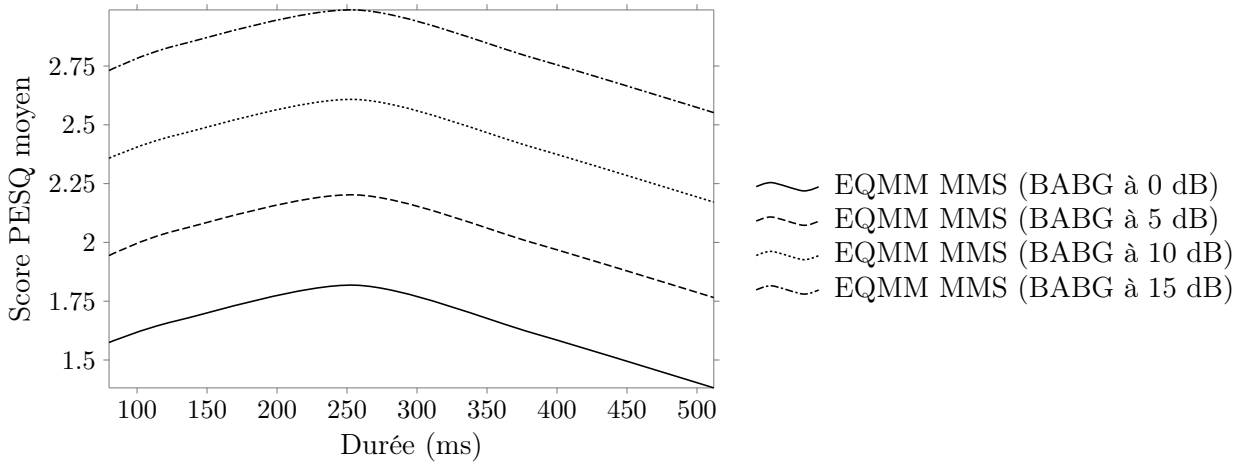

Figure 3.5 Effet de la durée d'une trame modulatoire sur le score PESQ pour l'algorithme EQMM MMS. Une durée de trame de 256 ms permet d'avoir le meilleur score PESQ.

Pour la durée d'une trame modulatoire, la figure 3.5 indique un gain de performance maximal à 256 ms, ce qui est aussi la durée permettant d'avoir une performance maximale avec une fréquence d'échantillonnage de 8 kHz [29].

Facteur de lissage α

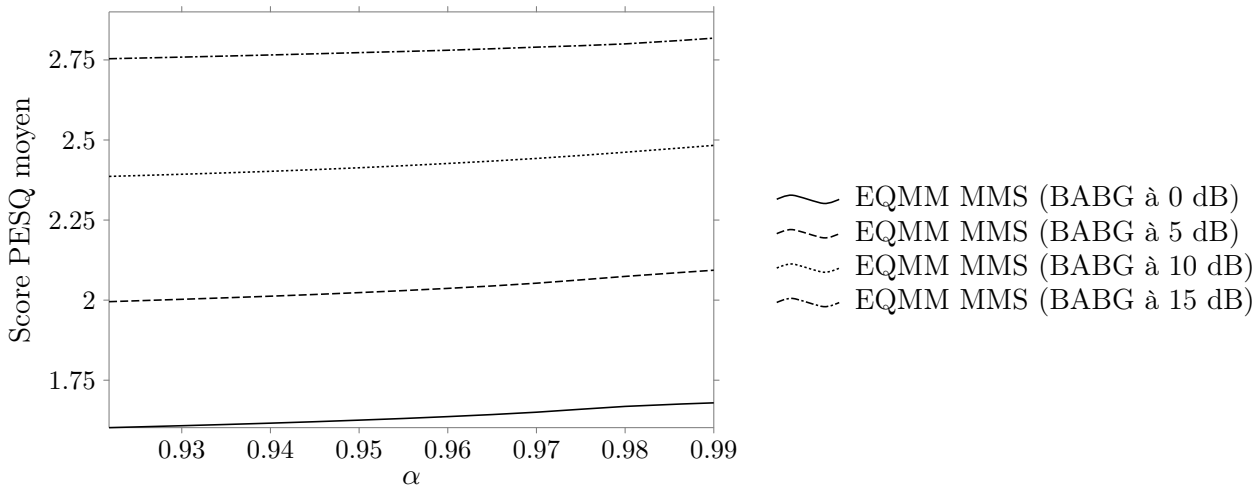

Figure 3.6 Effet du facteur de lissage α sur le score PESQ pour l'algorithme EQMM MMS. Une valeur de 0.99 permet d'obtenir le meilleur score PESQ.

La figure 3.6 indique aussi un niveau de performance maximum lorsque le facteur de lissage égale 0.99. La fréquence d'échantillonnage n'a donc pas d'incidence sur ce facteur dans le domaine des modulations du spectre.

Seuil minimum $\gamma_{\min}$ pour le RSB *a priori*

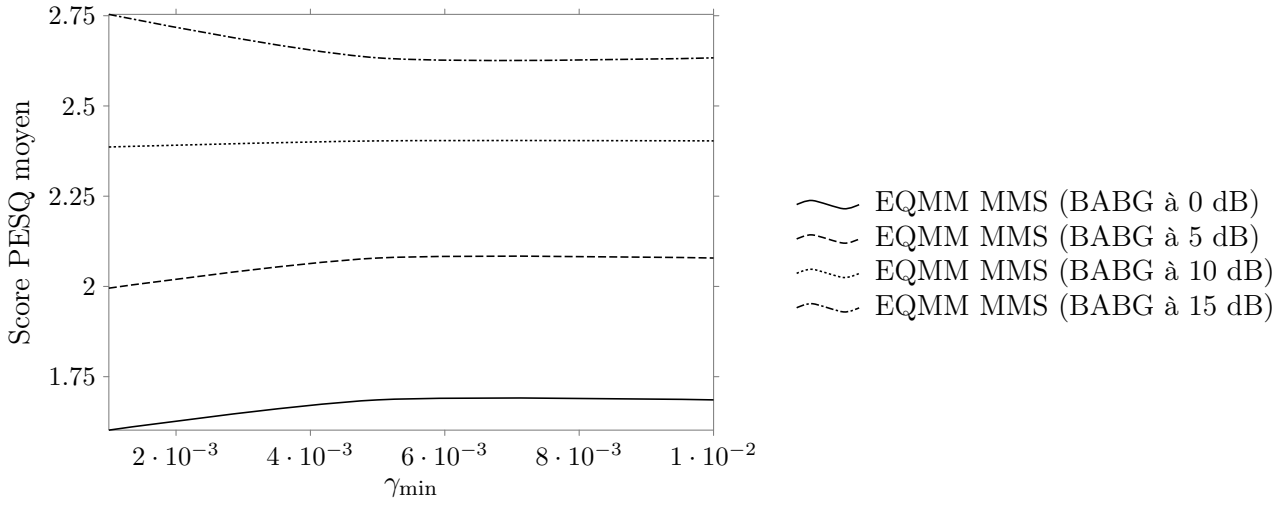

Figure 3.7 Effet du seuil minimum $\gamma_{\min}$ pour le RSB *a priori* sur le score PESQ pour l'algorithme EQMM MMS. Un seuil de 0.005 permet d'obtenir le score PESQ maximum à faible RSB.

Dans le domaine des modulations du spectre, la figure 3.7 indique qu'il n'y a pas de seuil minimum offrant le plus de performance pour tous les RSB étudiés. À 0.001, le score PESQ est à son maximum pour un RSB de 15 dB et à son pire pour les RSB 0 et 5 dB. À partir d'un seuil de 0.005, la valeur se stabilise jusqu'à 0.01. Le seuil de 0.005 est retenu pour avoir une meilleure performance à bas RSB, ce qui semble plus important que d'obtenir une meilleure performance à 15 dB de RSB.

Rehaussement par soustraction spectrale dans le domaine des modulations du spectre

Les paramètres étudiés pour l'estimateur par soustraction spectrale dans le domaine des modulations du spectre sont: longueur des trames spectrale, longueur des trames modulatoires, facteur de soustraction ρ et le seuil de soustraction β.

Longueur des trames spectrales

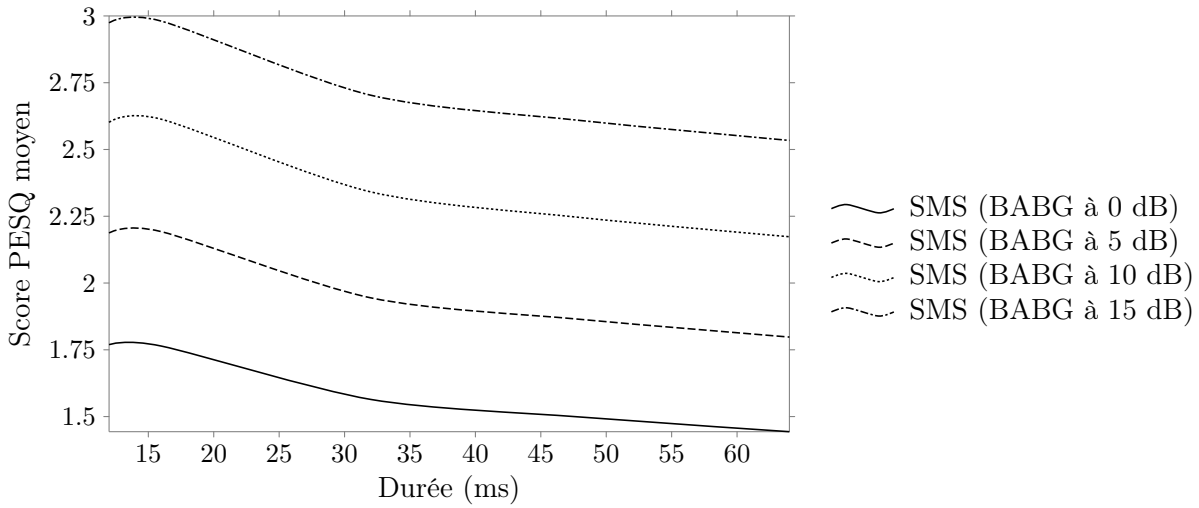

Figure 3.8 Effet de la durée d'une trame spectrale sur le score PESQ pour l'algorithme SMS. Une durée de trame de 12 ms permet d'obtenir le meilleur score PESQ.

La figure 3.8 indique une performance maximale de l'algorithme SMS lorsque la durée d'une trame est de 12 ms, soit un peu moins que le 16 ms obtenu pour les deux autres algorithmes. Par contre, la performance à 16 ms reste très bonne. Cela conclut donc qu'une longueur de trame de 16 ms va offrir les meilleures performances pour les trois algorithmes. Les travaux présentés en [29] [28] utilisaient plutôt des longueurs de trames de 32 ms. Le fait d'avoir doublé la fréquence d'échantillonnage a permis de réduire de moitié la longueur des trames spectrales.

Longueur des trames modulatoires

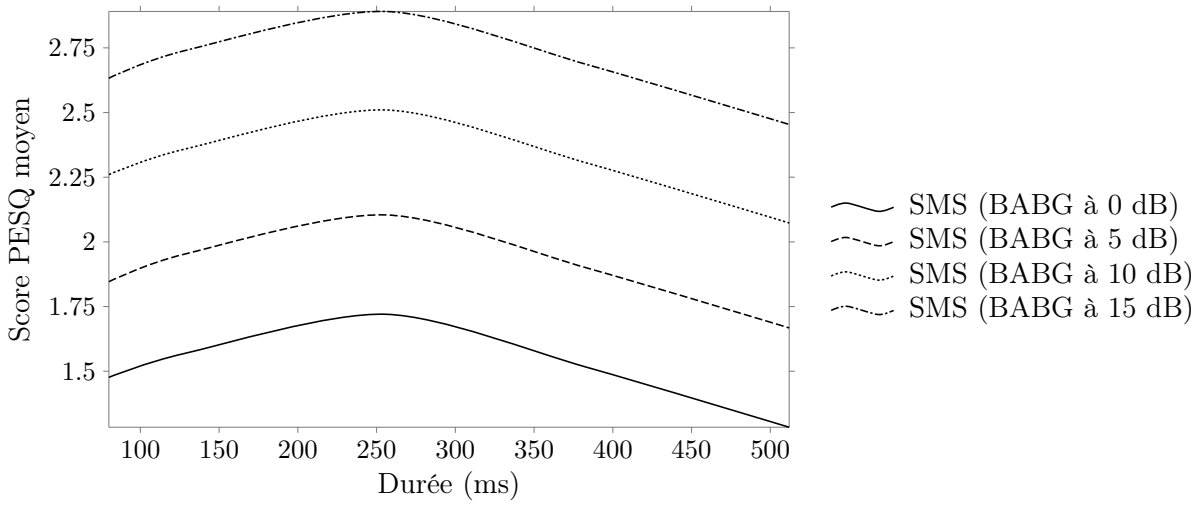

Figure 3.9 Effet de la durée d'une trame modulatoire sur le score PESQ pour l'algorithme SMS. Une durée de trame de 256 ms permet d'obtenir le meilleur score PESQ.

Pour le technique SMS, la figure 3.9 indique une durée de trame optimale de 256 ms, comme pour la technique EQMM MMS à 16 et 8 kHz. Ainsi, une durée de trame de 256 ms sera conservée pour la technique EQMM double.

Facteur de sursoustraction ρ

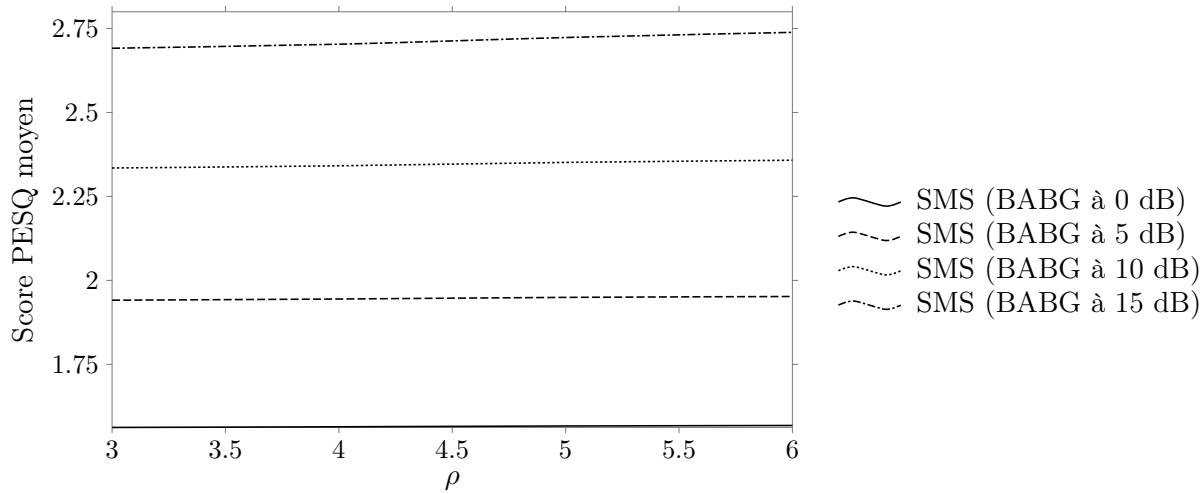

Figure 3.10 Effet du facteur ρ sur le score PESQ pour l'algorithme SMS. La variation du facteur n'a pas assez d'incidence sur le score PESQ. La valeur de 3 proposée par défaut est utilisée.

Pour le facteur de soustraction ρ, il y a une très faible amélioration pour les RSB de 10 et 15 dB lorsqu'il tend vers 6, une amélioration un peu plus notable à 15 dB et aucune différence à 0 dB. Les travaux présentés en [5] indiquaient que ce paramètre devait rester entre 3 et 6, avec 3 comme meilleure valeur la plupart du temps. Dans notre cas, la différence n'est pas assez notable pour changer ce paramètre. La valeur de ρ restera donc à 3.

Seuil de soustraction β

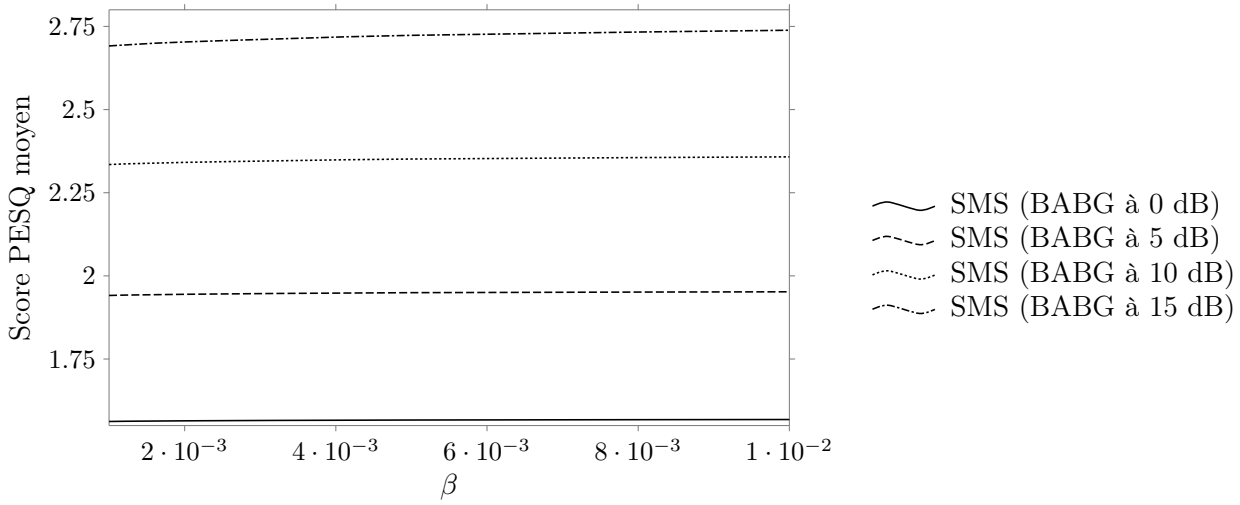

Figure 3.11 Effet du facteur β sur le score PESQ pour l'algorithme SMS. La valeur de 0.01 est utilisée étant donné le léger gain à 15 dB.

Le paramètre β, qui permet d'établir un seuil de soustraction, ne fait pas varier énormément la performance des algorithmes tels qu'indique la figure 3.11. Les performances pour les RSB de 0 et 5 dB restent identiques. Pour 10 dB, il y a une légère amélioration tandis qu'il y a une amélioration notable à 15 dB. Dans les travaux de [29], ce paramètre est fixé à 0.002. Étant donné le gain à 15 dB, il est modifié pour être à 0.01.

3.3 Développement de la fonction de combinaison de l'algorithme EQMM double

Cette section présente les modifications faites à la fonction de Fusion [29] (équation 2.43) afin d'utiliser un estimateur EQMM autant dans le domaine spectre que le domaine des modulations du spectre. Étant donné que seule la fonction dans le domaine des modulations du spectre a changé, seule la limite minimale a changé. À partir de ce changement,

l'équation (2.44) devient

$$\Psi_{\text{EQMM double}}\big[\sigma[l]\big] = \begin{cases} 0 & \text{si } g\big[\sigma[l]\big] \leq 5 \\ \frac{g\left[\sigma[l]\right]-5}{11} & \text{si } 5 < g\big[\sigma[l]\big] < 16 \\ 1 & \text{si } g\big[\sigma[l]\big] \geq 16 \end{cases} \cdot \tag{3.1}$$

La nouvelle limite de 5 dB, au lieu de 2 dB dans les travaux de [29], a été déterminée de manière empirique, en faisait varier cette limite. Le gain maximum de performance a été atteint autour de cette valeur. La limite supérieure a aussi été modifiée pour voir si un gain était possible de ce côté, mais la valeur de 16 dB reste la meilleure, ce qui reste logique étant donné que l'algorithme dans le domaine spectral n'a pas été modifié. L'équation (2.43) devient donc

$$\left|\hat{S}[\eta,l]\right| = \left(\Psi_{\text{EQMM double}}[\omega[l]]\left|S_{\text{EQMM AS}}[\eta,l]\right|^{\lambda} + (1-\Psi_{\text{EQMM double}}[\omega[l]])\left|S_{\text{EQMM MMS}}[\eta,l]\right|^{\lambda}\right)^{\frac{1}{\lambda}}. \tag{3.2}$$

Elle représente l'algorithme EQMM double dont les performances seront évaluées au chapitre suivant.

CHAPITRE 4

Résultats expérimentaux

Ce chapitre présente les résultats de rehaussement obtenu par l'algorithme proposé (EQMM double) et comparé aux algorithmes suivants: EQMM AS, Fusion, SMS et EQMM MMS. Trois mesures de performances objectives sont utilisées afin de comparer les performances des algorithmes soient le PESQ (*Perceptual Evaluation of Speech Quality*), le RSB par segments temporels et le RSD.

4.1 Mesure des performances d'un algorithme de rehaussement de la parole

Dans cette section, les trois mesures objectives de performance utilisées pour comparer les algorithmes sont présentées.

4.1.1 Mesure de qualité de la parole pour un système de télécommunication: PESQ

Le PESQ est un outil permettant d'offrir une méthodologie de tests permettant d'évaluer la qualité de la parole dans un contexte de télécommunication [17]. Il s'agit d'un outil officiellement utilisé par l'ITU-T, qui est l'organisme international s'assurant de la production des standards utilisés dans le domaine des télécommunications. Le but de cet outil est de pouvoir mesurer la qualité de la parole telle qu'évaluée par un humain. Il utilise un modèle se basant sur les tests subjectifs utilisés en télécommunication (comme le ITU-T P.800 [35]).
L'outil PESQ utilise un signal de référence, qui est le signal de parole non bruité dans notre cas, et le compare avec le signal à valider. Le résultat fourni par l'outil PESQ calque le principe de note d'opinion moyenne (MOS ou *Mean Opinion Score* en anglais), qui est utilisé dans les tests d'écoute subjectifs. Le score PESQ va de 1, qui signifie mauvais, à 5, qui signifie excellent.

4.1.2 RSB par segments temporels

Le RSB par segments temporels ou RSB segmental prend la moyenne du RSB de chaque trame utilisée dans le rehaussement, au lieu de prendre le RSB du signal au complet.

Comme le rehaussement se fait sur une trame à la fois, cet indicateur permet de disposer d'une mesure de performance directe de la capacité d'un algorithme de réduire le bruit dans une trame. L'équation est donnée par [21]

$$RSB_{seg} = \frac{10}{M} \sum_{m=0}^{M-1} \log_{10} \frac{\sum_{n=Lm}^{Lm+L-1} x^2[n]}{\sum_{n=Lm}^{Lm+L-1} (x[n] - \hat{x}[n])^2}. \tag{4.1}$$

où $x[n]$ est le signal de parole de référence (non bruité), $\hat{x}[n]$ est le signal de parole rehaussé, M est le nombre de trames désiré et L est la longueur de trame désirée. Plus la valeur du RSB est élevée, meilleur est le rehaussement.

4.1.3 Mesure de distorsion: Rapport signal à distorsion (RSD)

La dernière mesure de performance est le RSD, qui permet d'évaluer le niveau de distorsion entre le signal de parole rehaussé et le signal de parole de référence (sans bruit). Pour trouver cette mesure, le signal rehaussé est initialement décomposé en quatre parties à l'aide d'un filtre multicanal à temps invariant: le signal de référence, la distorsion, l'interférence et les artéfacts [37]. À partir de ces quatre parties, il est possible d'obtenir le RSD, qui est défini par

$$\mathrm{RSD} = 10 \log_{10} \frac{\left| s_{\text{réf}}[n] \right|^2}{\left| e_{\text{interf}}[n] + e_{\text{bruit}}[n] + e_{\text{artif}}[n] \right|^2} \tag{4.2}$$

où $s_{\text{réf}}[n]$ est le signal de parole de référence, $e_{\text{interf}}[n]$ le terme d'erreur d'interférence, e_{bruit} le terme d'erreur de bruit et $e_{\text{artif}}[n]$ le terme d'erreur d'artéfacts du signal de parole rehaussé. Cette mesure permet de connaître le niveau de distorsion induit dans le signal de parole lors du rehaussement. Un RSD plus élevé indique un niveau de distorsion plus faible.

4.2 Résultats et discussions

Les résultats sont regroupés selon la nature du bruit corrompant le signal soit premièrement, les bruits stationnaires et deuxièmement, les bruits non stationnaires.

4.2.1 Bruits stationnaires

Dans cette section, les résultats pour les bruits stationnaires sont présentés soient le bruit additif blanc gaussien et le bruit additif rose.

Bruit additif blanc gaussien (BABG)

Le BABG est un modèle statistique de bruit utilisé en théorie de l'information. Beaucoup de phénomènes naturels peuvent être représentés par ce modèle, ce qui en fait une bonne référence pour mesurer la performance des algorithmes de rehaussement. Les résultats PESQ du tableau 4.1 indiquent qu'à bas RSB, l'algorithme EQMM double, présenté dans la recherche, n'arrive pas à atteindre la performance des algorithmes EQMM dans les domaines spectral et des modulations du spectre. Par contre, à haut RSB, le principe d'utiliser les domaines spectral et des modulations du spectre permet d'avoir la meilleure performance. De plus, l'algorithme EQMM double surpasse légèrement l'algorithme de Fusion pour tous les niveaux de RSB.

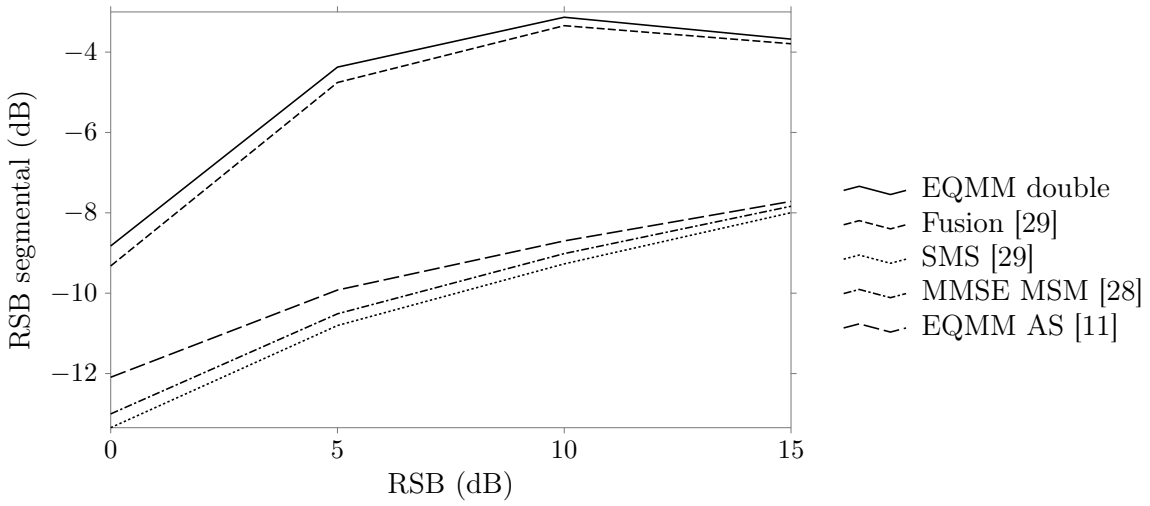

Figure 4.1 RSB segmental (dB) pour de la parole bruitée avec un BABG. L'algorithme EQMM double a le meilleur RSB segmental pour tout RSB.

Les résultats de RSB segmental de la figure 4.1 indiquent une grande amélioration de performance en utilisant deux domaines pour le rehaussement, tout RSB confondu. Cela démontre la capacité de l'algorithme EQMM double à trouver le meilleur rehaussement pour chaque trame. Ici encore, notre algorithme a une meilleure performance que l'algorithme de Fusion.

Tableau 4.1 Score PESQ moyen avec bruit additif blanc gaussien (BABG).

RSB (dB)	EQMM AS	Fusion	SMS	EQMM MMS	EQMM double
0	1.86	1.75	1.76	1.80	1.77
5	2.24	2.19	2.19	2.21	2.21
10	2.65	2.61	2.61	2.63	2.63
15	2.98	3.00	2.96	2.99	3.02

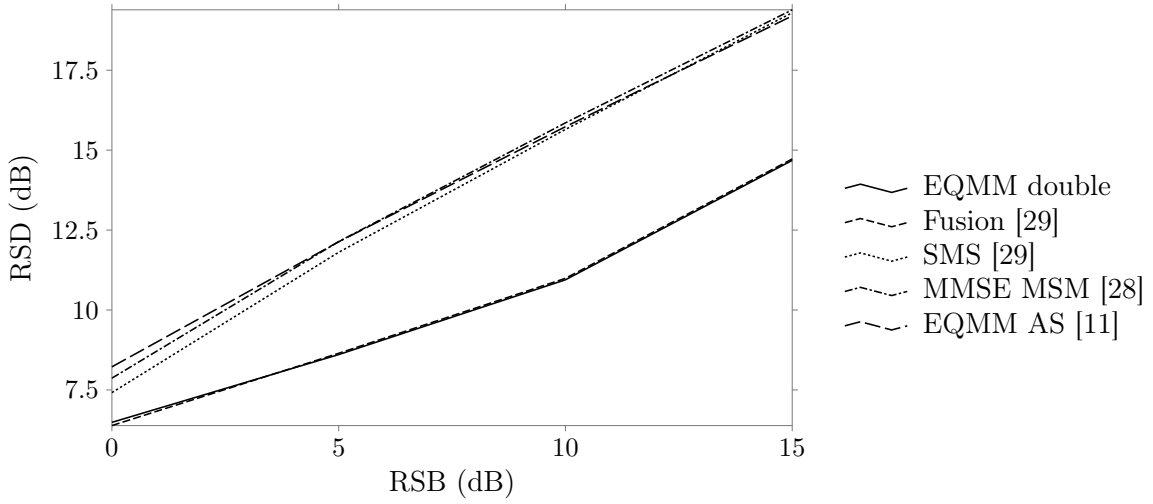

Figure 4.2 RSD (dB) pour de la parole bruitée avec un BABG. L'algorithme EQMM double présente plus de distorsion que les algorithmes à un seul domaine.

Bien que l'algorithme EQMM double réduit beaucoup plus l'énergie du bruit, il a le désavantage d'introduire de la distorsion, tel que vu par les résultats de la figure 4.2. Les deux techniques utilisant les deux domaines ont de moins bons résultats que ceux utilisant uniquement un seul domaine. La technique EQMM double et Fusion ont quasiment le même niveau de distorsion. C'est l'algorithme EQMM AS qui génère le moins de distorsion à faible RSB, tandis que tous les algorithmes à simple domaine ont le même niveau de distorsion à fort RSB.

Bruit additif rose

Le bruit additif rose (BAR) est un bruit dont la DSP est inversement proportionnelle à la fréquence du signal. Le terme rose vient du fait que le spectre de la lumière visible ayant cette DSP apparait comme rose [20]. Il s'agit d'un bruit dont la DSP apparait dans plusieurs phénomènes naturels [12] et est considérée comme omniprésente [2]. Il est donc pertinent de tester un algorithme de rehaussement de la parole avec un bruit rose étant donné son omniprésence. Pour le score PESQ du tableau 4.2 avec un BAR, l'algorithme

Tableau 4.2 Score PESQ moyen avec bruit additif rose (BAR).

RSB (dB)	EQMM AS	Fusion	SMS	EQMM MMS	EQMM double
0	1.96	1.89	1.87	1.91	1.92
5	2.38	2.33	2.32	2.35	2.36
10	2.70	2.71	2.69	2.71	2.73
15	3.04	3.09	3.04	3.05	3.09

EQMM double a une meilleure performance qu'avec un BABG. Seul l'algorithme EQMM

AS a une meilleure performance que lui à faible RSB. À haut RSB, son résultat est équivalent à la technique de Fusion.

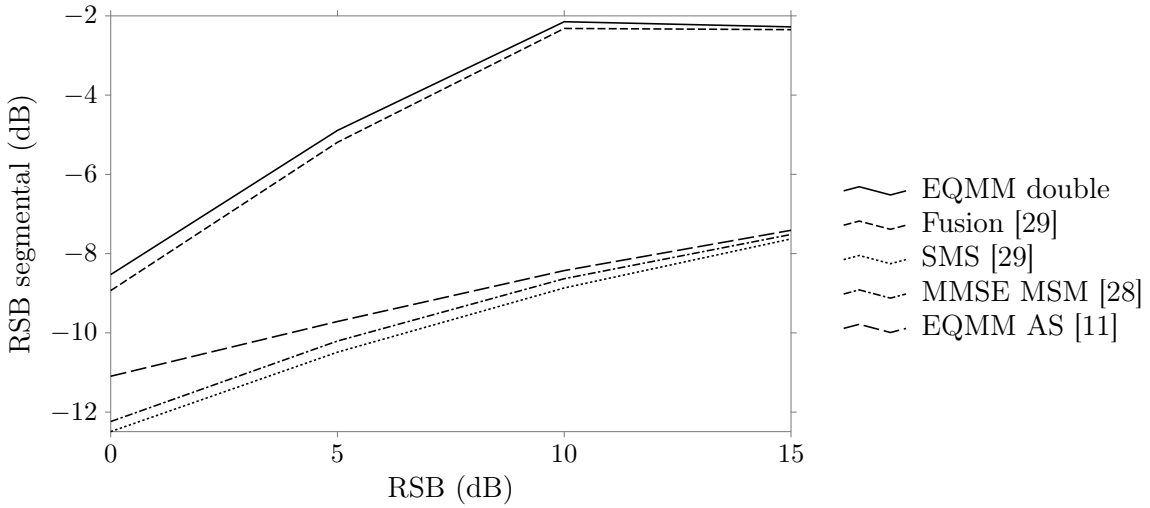

Figure 4.3 RSB segmental (dB) pour de la parole bruitée avec un BAR. L'algorithme EQMM double a le meilleur RSB segmental pour tout RSB.

Pour le RSB segmental de la figure 4.3, la technique EQMM double surpasse les techniques avec seulement un domaine pour tout niveau de RSB. Elle est aussi légèrement supérieure à la technique de Fusion. Les résultats sont semblables à ceux avec un BABG, ce qui correspond à ce qui était attendu étant donné la similitude entre les deux types de bruit. Le RSB segmental à 15 dB est légèrement plus faible qu'à 10 dB pour la technique EQMM double. Cette anomalie n'est pas reproduite dans le score PESQ ou dans le RSD.

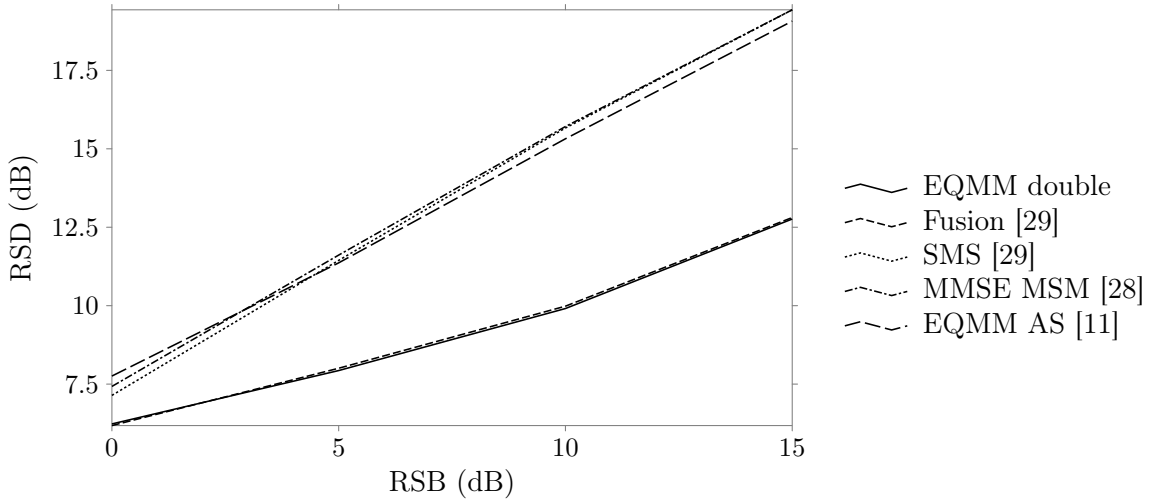

Figure 4.4 RSD (dB) pour de la parole bruitée avec un BAR. L'algorithme EQMM double présente plus de distorsion que les algorithmes à un seul domaine.

Selon la figure 4.4, les deux techniques utilisant les deux domaines pour le rehaussement ont le même problème de distorsions qu'avec un BABG. Les techniques de Fusion et EQMM double ont la même performance. La technique EQMM AS est la meilleure avec un RSB de 0 dB, mais devient la moins performante à fort RSB pour les techniques avec un seul domaine.

4.2.2 Bruits non stationnaires

Dans cette section, les résultats pour les bruits non stationnaires de conversation, d'usine et de voiture sont présentés.

Bruit de conversation

Le bruit de conversation provient de conversations se déroulant dans un environnement de bureau [36]. Pour ce premier bruit non stationnaire, le score PESQ du tableau 4.3 est

Tableau 4.3 Score PESQ moyen avec bruit additif de conversation.

RSB (dB)	EQMM AS	Fusion	SMS	EQMM MMS	EQMM double
0	1.74	1.67	1.68	1.71	1.70
5	2.11	2.10	2.08	2.11	2.11
10	2.53	2.56	2.56	2.58	2.56
15	2.81	2.84	2.80	2.83	2.84

plus faible par rapport aux bruits stationnaires pour toutes les techniques et tous les RSB, ce qui est normal étant donné qu'il est plus difficile d'estimer le bruit. À faible RSB, la

technique EQMM double est moins performante que EQMM AS et EQMM MMS. À fort RSB, les techniques EQMM double et Fusion sont les plus performantes, avec EQMM MMS ayant un score presque équivalent.

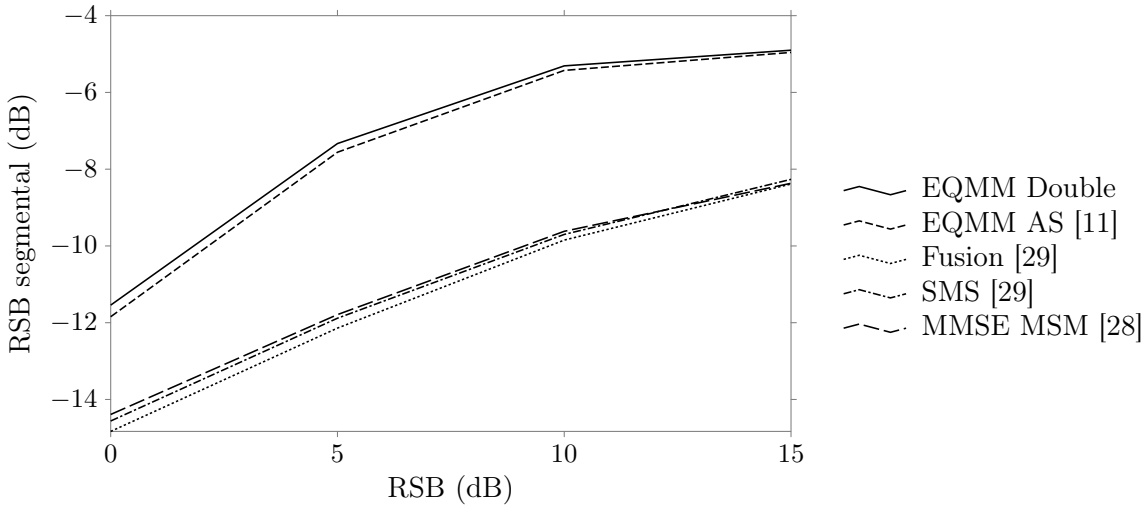

Figure 4.5 RSB segmental (dB) pour de la parole bruitée avec bruit additif de conversations. L'algorithme EQMM double a le meilleur RSB segmental pour tout RSB.

Le RSB segmental de la figure 4.5 est aussi dans une plage plus faible par rapport au RSB segmental pour un BABG et BAR. Les algorithmes ont une performance semblable par rapport aux autres. Les techniques de Fusion et EQMM double restent les plus performantes, avec la technique EQMM double offrant de meilleures performances.

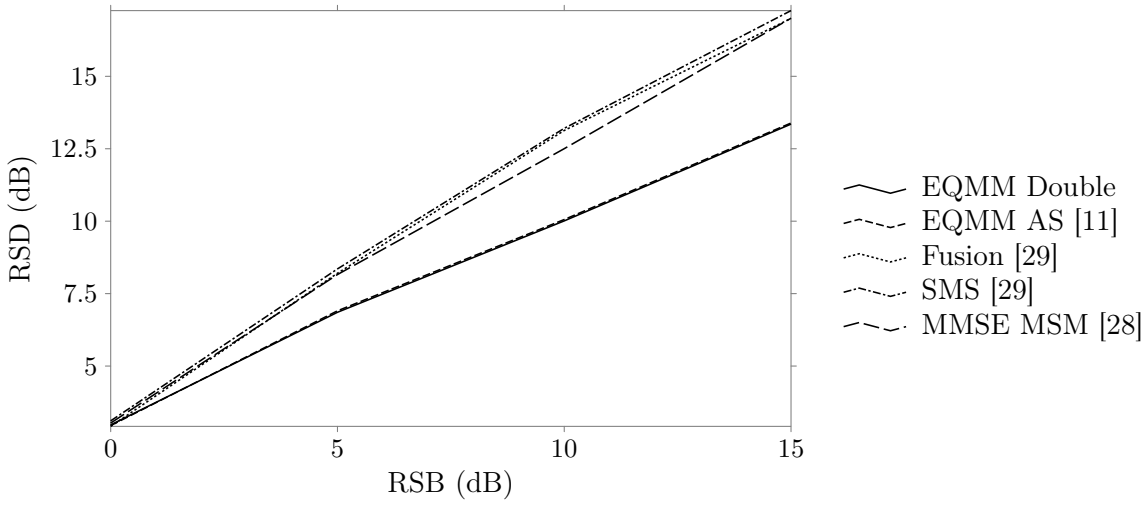

Figure 4.6 RSD (dB) pour de la parole bruitée avec bruit additif de conversations. L'algorithme EQMM double présente plus de distorsion que les algorithmes à un seul domaine.

Le RSD de la figure 4.6 pour l'algorithme EQMM double est plus positif pour ce premier bruit non stationnaire. À 0 dB, tous les algorithmes ont un même niveau de distorsions, alors que la technique EQMM double a un meilleur score PESQ et un meilleur RSB segmental. À plus fort RSB, les techniques EQMM double et Fusion n'arrivent pas à avoir un RSD du même niveau que le reste des algorithmes.

Bruit d'usine

Tableau 4.4 Score PESQ moyen avec bruit additif d'usine.

RSB (dB)	EQMM AS	Fusion	SMS	EQMM MMS	EQMM double
0	1.76	1.69	1.70	1.74	1.72
5	2.15	2.11	2.11	2.14	2.13
10	2.54	2.54	2.52	2.55	2.56
15	2.90	2.93	2.89	2.92	2.94

Le score PESQ du tableau 4.4 pour de la parole en présence de bruits d'usine garde la même structure que pour les scores précédents. À faible RSB, la technique développée dans ce projet de recherche arrive en troisième place, après EQMM AS et EQMM MMS, mais a le meilleur score à fort RSB. Elle a par contre de meilleures performances que la technique de Fusion pour tout RSB.

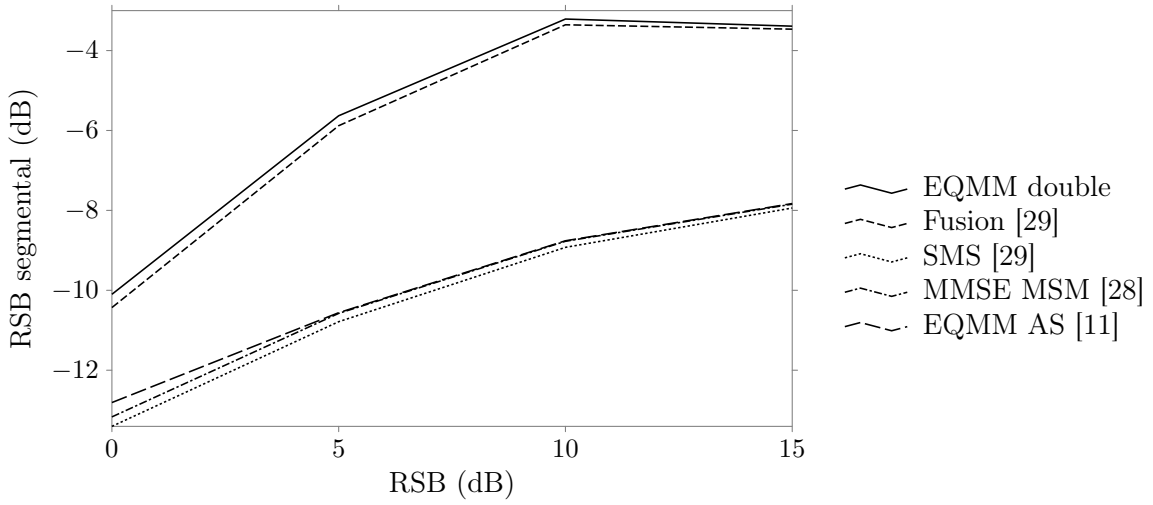

Figure 4.7 RSB segmental (dB) pour de la parole bruitée avec un bruit additif d'usine. L'algorithme EQMM double a le meilleur RSB segmental pour tout RSB.

Les techniques EQMM double et Fusion sont encore les plus performantes pour le RSB segmental de la figure 4.7. Comme les résultats avec un BAG, le RSB segmental à 15 dB est légèrement plus faible qu'à 10 dB pour les techniques EQMM double de Fusion.

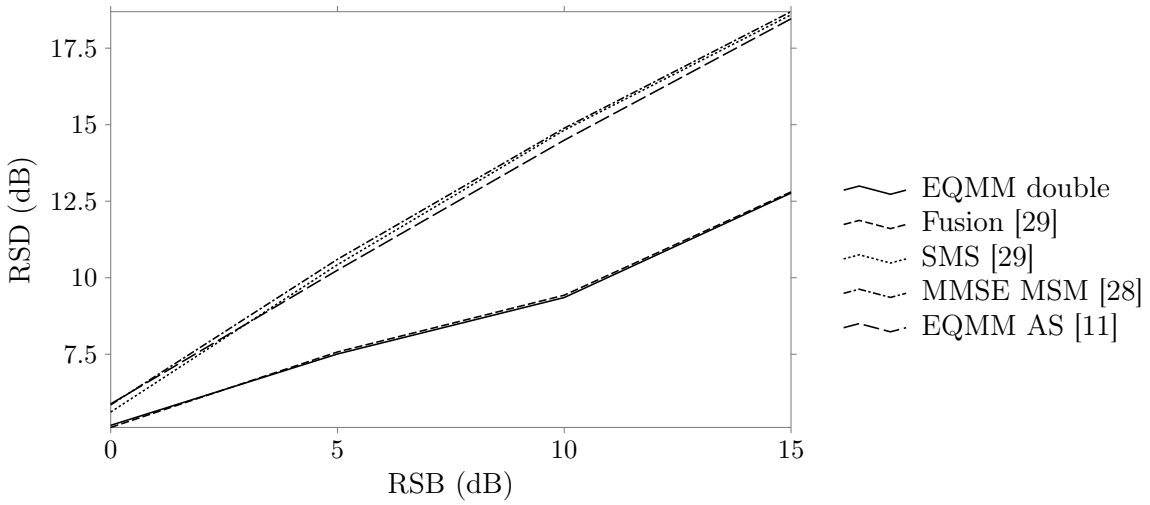

Figure 4.8 RSD (dB) pour de la parole bruitée avec un bruit additif d'usine. L'algorithme EQMM double présente plus de distorsion que les algorithmes à un seul domaine.

Pour le RSD de la figure 4.8, les résultats restent semblables à faible RSB, avec un petit avantage pour les techniques à un seul domaine. Les techniques EQMM double et Fusion n'arrivent pas par contre à un niveau de distorsions aussi faible que les techniques à un seul domaine.

Bruit de voiture

Le bruit de voiture est particulier, car il est présent seulement dans les basses fréquences et masque très peu la parole. Il est possible d'entendre sans aucun problème le texte prononcé même à 0 dB, alors qu'il est très difficile de le comprendre pour les autres types de bruit. Cela a comme conséquence d'avoir de très bonne performance en général pour les trois mesures de performance. Malgré cela, l'environnement intérieur d'un véhicule reste pertinent à utiliser étant donné qu'il s'agit d'un endroit où des communications téléphoniques ont lieu. Le tableau 4.5 indique que la technique EQMM double a la meilleure performance à 15 dB, mais diminue pour avoir la moins bonne performance à 0 dB.

Tableau 4.5 Score PESQ moyen avec bruit additif d'usine.

RSB (dB)	EQMM AS	Fusion	SMS	EQMM MMS	EQMM double
0	3.34	3.34	3.30	3.32	3.28
5	3.68	3.68	3.59	3.60	3.63
10	3.86	3.85	3.75	3.79	3.85
15	4.02	4.00	3.87	3.92	4.03

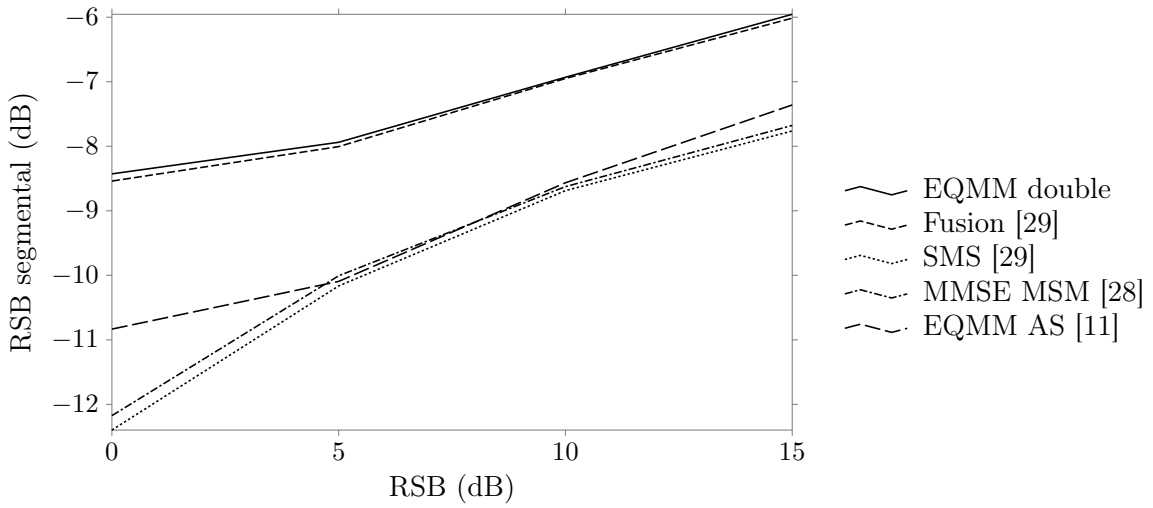

Figure 4.9 RSB segmental (dB) pour de la parole bruitée avec un bruit additif de voiture. L'algorithme EQMM double a le meilleur RSB segmental pour tout RSB.

Pour le RSB segmental, la figure 4.9 indique que les deux techniques à deux domaines ont des résultats supérieurs aux autres techniques, avec la technique EQMM double légèrement supérieure à la technique de Fusion.

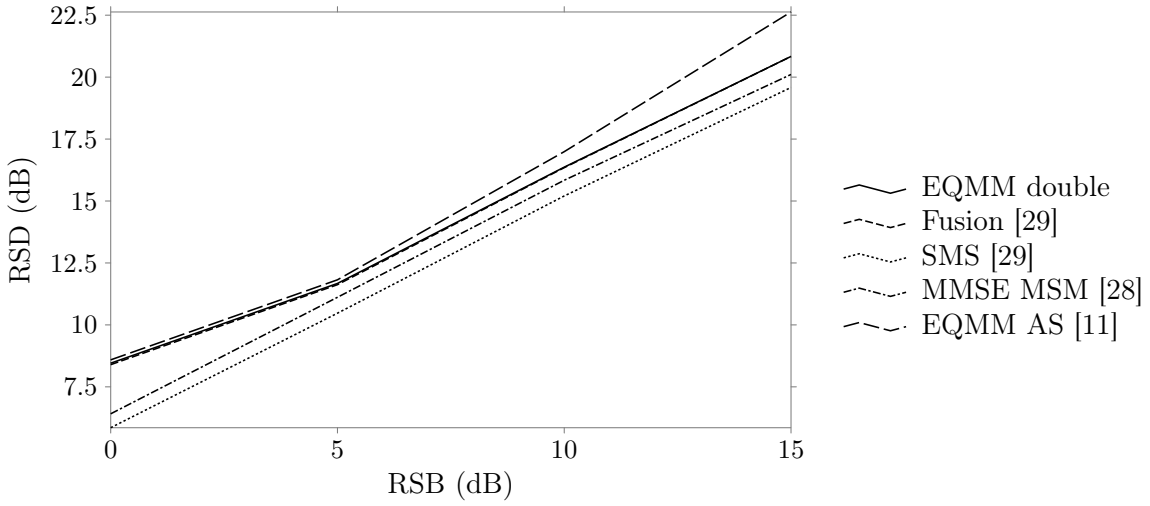

Figure 4.10 RSD (dB) pour de la parole bruitée avec un bruit additif de voiture. L'algorithme EQMM double présente plus de distorsion que les algorithmes à un seul domaine.

Finalement, pour la dernière mesure, la figure 4.10 indique, pour la première fois, un niveau de RSD supérieur pour les techniques à deux domaines. Les techniques EQMM double et Fusion ont une performance similaire à 0 et 5 dB, tandis que la technique de Fusion a de meilleurs résultats à 10 et 15 dB, suivit de près par la technique de EQMM double.

CHAPITRE 5

Conclusion

Ce projet de recherche visait à développer un algorithme de rehaussement de la parole utilisant les domaines spectral et des modulations du spectre et opérant à une fréquence d'échantillonnage de 16 kHz. Comme vu initialement, les systèmes de traitement de parole fonctionnant en environnement non contrôlé opèrent avec des signaux de parole bruités. Il est donc nécessaire pour leur bon fonctionnement de réduire au maximum la présence de bruit.

L'approche proposée, EQMM double, montre de meilleurs résultats pour le RSB segmental que tous les autres algorithmes comparés, et ce pour tous les types de bruits étudiés. De plus, en termes de PESQ, il semble être le meilleur estimateur pour les bruits stationnaires à haut RSB et est comparable aux autres algorithmes pour les autres conditions. Pour ce qui est du RSD, les deux méthodes de fusion semblent introduire davantage de distorsion que les autres méthodes comparées. Le PESQ est généralement davantage corrélé avec les distorsions alors que le RSB segmental est davantage corrélé avec la réduction de bruit [16]. Cela semble donc indiquer que l'algorithme EQMM double réduit davantage le bruit que les autres algorithmes comparés, mais en contrepartie il introduit également davantage de distorsion. Ce compromis entre la réduction du bruit et l'ajout de distorsion dans la parole est bien connu dans le domaine du rehaussement de la parole.

L'utilisation d'une fréquence d'échantillonnage de 16 kHz a également été validée en étudiant l'effet de ce changement de cadence sur les différents paramètres des algorithmes étudiés. À l'exception de la longueur de trame, qui est directement liée à la fréquence d'échantillonnage, les autres paramètres ont conservé essentiellement les valeurs utilisées à 8 kHz.

En terminant, voici quelques pistes qui pourraient permettre d'améliorer les performances de l'algorithme proposé:

1. *Amélioration de la fonction de combinaison avec prise de décision a priori.* Dans sa version actuelle, l'algorithme rehausse le signal de parole dans le domaine spectral, puis dans le domaine des modulations du spectre, pour finalement prendre une décision sur la combinaison de chaque domaine à utiliser. Il pourrait être per-

tinent de trouver une manière de déterminer avant le rehaussement quel domaine ou combinaison de domaines seront utilisés. Cette amélioration aura un impact sur la troisième piste d'amélioration ci-dessous, étant donné qu'il sera possible d'éviter certains calculs si seulement un domaine est nécessaire.

2. *Explorer d'autres types de combinaison.* La fonction de combinaison utilisée est linéaire. La parole étant un signal complexe, des fonctions de combinaison plus complexes ou des fonctions indépendantes pour des plages de fréquences données pourraient offrir de meilleures performances.

3. *Optimisation pour temps réel.* L'utilisation du domaine des modulations du spectre fait augmenter de façon exponentielle le nombre de calculs. Il serait intéressant de trouver des techniques d'optimisation réduisant le temps de calcul pour permettre son fonctionnement en temps réel sur des systèmes de type consommateur (p. ex. téléphone cellulaire).

LISTE DES RÉFÉRENCES

[1] Atlas, L. et Shamma, S. A. (2003). Joint acoustic and modulation frequency. *EURASIP Journal on Advances in Signal Processing*, volume 2003, numéro 7, p. 1–8.

[2] Bak, P., Tang, C. et Wiesenfeld, K. (1987). Self-organized criticality: An explanation of the 1/f noise. *Physical review letters*, volume 59, numéro 4, p. 381.

[3] Behar, R. (2017). *Everything you need to know about VoLTE in Canada.* `https://mobilesyrup.com/2017/11/13/canada-volte-guide/` (page consultée le 2018.03.24).

[4] Benesty, J., M. M. S. et Huang, Y. (2008). *Springer handbook of speech processing.* Springer Berlin Heidelberg.

[5] Berouti, M., Schwartz, R. et Makhoul, J. (1979). Enhancement of speech corrupted by acoustic noise. Dans *IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 4. p. 208–211.

[6] Boll, S. (1979). Suppression of acoustic noise in speech using spectral subtraction. *IEEE Transactions on Acoustics, Speech and Signal Processing*, volume 27, numéro 2, p. 113–120.

[7] Bosco, J. et Plourde, E. (2017). Speech enhancement using both spectral and spectral modulation domains. Dans *2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE)*. p. 1–4.

[8] Brillinger, D. (1975). *Time Series: Data Analysis and Theory.* Numéro v. 1 dans International Series in Decision Processes, Holt, Rinehart, and Winston.

[9] Cappe, O. (1994). Elimination of the musical noise phenomenon with the ephraim and malah noise suppressor. *IEEE Transactions on Acoustics, Speech and Signal Processing*, volume 2, numéro 2, p. 345–349.

[10] E. H. Rothauser, W. D. C. (1969). Ieee recommended practice for speech quality measurements. *IEEE No 297-1969*, p. 1–24.

[11] Ephraim, Y. et Malah, D. (1984). Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator. *IEEE Transactions on Acoustics, Speech and Signal Processing*, volume 32, numéro 6, p. 1109–1121.

[12] Gardner, M. (1970). Mathematical games. *Scientific American*, volume 222, numéro 6, p. 132–140.

[13] GSM Association (2016). Ims profile for voice and sms. *GSMA*. `https://www.gsma.com/newsroom/wp-content/uploads//IR.92-v10.0.pdf` (page consultée le 2018.03.24).

[14] Harris, F. J. (1978). On the use of windows for harmonic analysis with the discrete fourier transform. *Proceedings of the IEEE*, volume 66, numéro 1, p. 51–83.

[15] Hirsch, H.-G. et Ehrlicher, C. (1995). Noise estimation techniques for robust speech recognition. Dans *Acoustics, Speech, and Signal Processing, 1995. ICASSP-95., 1995 International Conference on*, IEEE. Volume 1. p. 153–156.

[16] Hu, Y. et Loizou, P. C. (2008). Evaluation of objective quality measures for speech enhancement. *IEEE Transactions on audio, speech, and language processing*, volume 16, numéro 1, p. 229–238.

[17] ITU-T Recommendation P.862 (2001). Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs.

[18] ITU-T Study Groups 12, 15 and 16 (2010). G.191 : Software tools for speech and audio coding standardization. *International Telecommunication Union*. `https://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-G.191-201003-I!!SOFT-ZST-E&type=items` (page consultée le 2018.03.24).

[19] Kabal, P. (2002). Tsp speech database. *Telecommunications & Signal Processing Laboratory*. `http://www-mmsp.ece.mcgill.ca/Documents/Downloads/TSPspeech/TSPspeech.pdf` (page consultée le 2018.03.24).

[20] Keshner, M. S. (1982). 1/f noise. *Proceedings of the IEEE*, volume 70, numéro 3, p. 212–218.

[21] Kondo, K. (2012). *Subjective Quality Measurement of Speech: Its Evaluation, Estimation and Applications*. Signals and Communication Technology, Springer Berlin Heidelberg.

[22] Martin, R. (1994). Spectral subtraction based on minimum statistics. Dans *in Proc. Euro. Signal Processing Conf. (EUSIPCO)*. p. 1182–1185.

[23] Martin, R. (2001). Noise power spectral density estimation based on optimal smoothing and minimum statistics. *IEEE Transactions on Speech and Audio Processing*, volume 9, numéro 5, p. 504–512.

[24] McAulay, R. et Malpass, M. (1980). Speech enhancement using a soft-decision noise suppression filter. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, volume 28, numéro 2, p. 137–145.

[25] Meyer, J., Simmer, K. U. et Kammeyer, K. D. (1997). Comparison of one- and two-channel noise-estimation techniques. Dans *Proc. 5th International Workshop on Acoustic Echo and Noise Control, IWAENC-97*. p. 137–145.

[26] Oppenheim, A. et Schafer, R. (2011). *Discrete-Time Signal Processing*. Pearson Education.

[27] Paliwal, K., Schwerin, B. et Wojcicki, K. (2011). Role of modulation magnitude and phase spectrum towards speech intelligibility. *Speech Communication*, volume 53, numéro 3, p. 327 – 339.

[28] Paliwal, K., Schwerin, B. et Wojcicki, K. (2012). Speech enhancement using a minimum mean-square error short-time spectral modulation magnitude estimator. *Speech Communication*, volume 54, numéro 2, p. 282–305.

[29] Paliwal, K., Wojcicki, K. et Schwerin, B. (2010). Single-channel speech enhancement using spectral subtraction in the short-time modulation domain. *Speech Communication*, volume 52, numéro 5, p. 450–475.

[30] Raphael, L. (2011). *Speech science primer : physiology, acoustics, and perception of speech*. Wolters Kluwer Health/Lippincott Williams & Wilkins, Baltimore, MD.

[31] Reininger, A. (2012). *Wolfgang von Kempelen : a biography*. East European Monographs Distributed by Columbia University Press, Boulder, Colo. New York.

[32] Telecommunication standardization sector of ITU (1988). Modulation par impulsions et codage (mic) des frÉquences vocales. *International Telecommunication Union*. `https://www.itu.int/rec/T-REC-G.711-198811-I/fr` (page consultée le 2018.03.24).

[33] Telecommunication standardization sector of ITU (2011). P.56 : Mesure objective du niveau vocal actif. *International Telecommunication Union*. `http://www.itu.int/rec/T-REC-P.56-201112-I/fr` (page consultée le 2018.03.24).

[34] Telecommunication standardization sector of ITU (2012). G.711.1 : Extension intégrée large bande du système de modulation par impulsions et codage g.711. *International Telecommunication Union*. `https://www.itu.int/rec/T-REC-G.711.1-201209-I/fr` (page consultée le 2018.03.24).

[35] Telecommunication standardization sector of ITU (2012). P.800 : Méthodes d'évaluation subjective de la qualité de transmission. *International Telecommunication Union*. `https://www.itu.int/rec/T-REC-P.800-199608-I/fr` (page consultée le 2018.03.24).

[36] Varga, A. et Steeneken, H. J. (1993). Assessment for automatic speech recognition: Ii. noisex-92: A database and an experiment to study the effect of additive noise on speech recognition systems. *Speech communication*, volume 12, numéro 3, p. 247–251.

[37] Vincent, E., Gribonval, R. et Fevotte, C. (2006). Performance measurement in blind audio source separation. *IEEE Transactions on Audio, Speech, and Language Processing*, volume 14, numéro 4, p. 1462–1469.

[38] Zadeh, L. (1950). Frequency analysis of variable networks. *Proceedings of the IRE*, volume 38, numéro 3, p. 291–299.